MIT LIBRARIES DUPL
3 9080 02874 5161
HB31
.M415
10-09-14
2009
DEWEY

Massachusetts Institute of Technology
Department of Economics
Working Paper Series

Marry for What?
Caste and Mate Selection in Modern India.

Abhijit Banerjee
Esther Duflo
Maitreesh Ghatak
Jeanne Lafortune

Working Paper 09-14
May 8, 2009
Revised: May 19, 2009

Room E52-251
50 Memorial Drive
Cambridge, MA 02142

This paper can be downloaded without charge from the
Social Science Research Network Paper Collection at
http://ssrn.com/abstract=1401576

Marry for What? Caste and Mate Selection in Modern India*

Abhijit Banerjee, Esther Duflo, Maitreesh Ghatak and Jeanne Lafortune†

May 2009

Abstract

This paper studies the role played by caste, education and other social and economic attributes in arranged marriages among middle-class Indians. We use a unique data set on individuals who placed matrimonial advertisements in a major newspaper, the responses they received, how they ranked them, and the eventual matches. We estimate the preferences for caste, education, beauty, and other attributes. We then compute a set of stable matches, which we compare to the actual matches that we observe in the data. We find the stable matches to be quite similar to the actual matches, suggesting a relatively frictionless marriage market. One of our key empirical findings is that there is a very strong preference for within-caste marriage. However, because both sides of the market share this preference and because the groups are fairly homogeneous in terms of the distribution of other attributes, in equilibrium, the cost of wanting to marry within-caste is low. This allows caste to remain a persistent feature of the Indian marriage market.

*We thank the *Anandabazar Patrika* for their cooperation for this project, and Prasid Chakrabarty and the team of SRG investigators for conducting the survey. We thank Raquel Fernandez, Ali Hortascu, Patrick Bajari, George Mailath, Whitney Newey, Parag Pathak, Andrew Postelwaithe, Debraj Ray, Alvin Roth, Ken Wolpin and many seminar participants for very helpful comments. Finally, we also thank Sanchari Roy and Tommy Wang for outstanding research assistance.

†The authors are from the departments of economics at MIT, MIT, LSE, and University of Maryland, College Park respectively.

Digitized by the Internet Archive
in 2011 with funding from
Boston Library Consortium Member Libraries

http://www.archive.org/details/marryforwhatcast00bane

1 Introduction

Marriage is, among other things, an important economic decision. Sorting in families impacts children outcomes and log term inequality (Fernandez and Rogerson (2001), Fernandez (2003)). In developing countries, where many women do not work outside their homes, marriage is arguably the single most important determinant of a woman's economic future. In India, the setting for this study, several studies have shown that marriage is indeed taken as a very serious economic decision, managed by parents more often than by the prospective spouses. For example, Rosenzweig and Stark (1989) show that parents marry their daughters in villages where incomes co-vary less with respect to their own village. Foster and Rosenzweig (2001) show that demand for healthy women in the marriage market influences investments in girls.

Yet, despite the economic importance of this decision, "status"-like attributes, such as caste, continue to play a seemingly crucial role in determining marriage outcomes in India. In a recent opinion poll in India, 74 percent of respondents declared to be opposed to inter-caste marriage.[1] The institution is so prevalent that matrimonial advertisements (henceforth, ads) in Indian newspapers are classified under caste headings, making it immediately obvious where prospective brides or grooms can find someone from their own caste. This caste system, where offsprings of an inter-caste couple are supposed to lose their caste, resembles the "aristocratic equilibrium" of Cole et al. (1992), which, they show, is also characterized by low levels of productivity.[2]

Cole et al. (1992) go on to show that the aristocratic equilibrium may be broken by economic mobility when, for example, a high wealth, low status man may find it profitable to marry a high status, low wealth woman. Economic growth and the diversification of earnings opportunities have significantly lowered the correlation between caste and income in India, and the role of caste in job matching has declined (Munshi and Rosenzweig 2006). Will this lead to a decline in the role of caste in marriage decisions, as the constraints it imposes become too costly to be sustained in equilibrium?

To answer this question, we need to know whether the strong preference means that caste has a large effect on marital matching. Do people end up marrying someone very different (in terms of attributes other than caste) from those whom they would have married absent this regard for caste? Do we actually see the distortion in choices relative to what would be observed in a caste-free world?

This paper sheds light on these questions. First, a simple model is developed to characterize

[1] We use the word caste in the sense of *jati* (community) as opposed to *varna*. The latter is a broad theoretical system of grouping by occupation (priests, nobility, merchants, and workers). The jati is the community within which one is required to be married, and which forms one's social identity.

[2] The formal rule may be that the children of an inter-caste couple inherit the caste of the father, but in practice, they tend to be discriminated against in the conservative segments of society.

the marriage market equilibrium when individuals value caste as well as the more standard characteristics such as attractiveness or income. We characterize conditions under which inter-caste marriages may take place. We show that the influence of caste preferences on the marriage market equilibrium depends crucially on the type of preference over caste, and the distribution of non-caste attributes across the population. In the case where preferences for caste are primarily "horizontal" (people care more about marrying someone from the same caste than about marrying "up"), preference for within-caste marriage does not change the equilibrium matching patterns when the distribution of male and female (non-caste) attributes are balanced (the distribution of the "quality" of men within each caste resembles the distribution of the quality of women).[3] This will be true even if the "demand-price" of caste (how much people are willing to give up in terms of partner quality to marry within caste) is very high. The reason is that with horizontal preferences people prefer to marry in caste and by the balanced population assumption anyone they could realistically expect to marry outside their caste, has the option of matching with somebody who is very similar from within their own caste.

In contrast if caste is primarily vertical, then we show that the intensity of preference for within-caste marriage or marrying up in caste affects the entire pattern of who matches with whom in equilibrium. This will also be the case if the population is highly unbalanced, even with horizontal caste preferences because then even though people want to marry within caste, there may not be any suitable candidates available for them to do so.

The theoretical framework suggests that one of the key elements in determining the impact of caste preferences in marriage markets thus lies in the nature and intensity of these preferences. We analyze an unusual data set on the arranged marriage market that we collected in Kolkata, the capital of the state of West Bengal in India. We interviewed a random sample of 783 individuals who placed matrimonial ads (henceforth, ad-placers) in the major Bengali newspaper, *Anandabazar Patrika*, which, with its circulation of 1.2 million, is the largest circulated single edition daily newspaper (across all languages) in India.[4] Most ad-placers are parents who are placing an ad on behalf of their son or daughter.

The sample consists mainly of educated, relatively well-off, urban middle-class Bengalis: 85 percent of both the prospective grooms and brides have a college degree. This is of course a special population that would be a problem if our goal was to describe the preferences of the representative Indian. However, as explained above, our primary interest is more conceptual – we want to understand the relationship between the nature of individual preferences and the

[3] In other words it is not, for example, the case that all the women from one caste are at the 90th percentile of the population distribution in terms of say, attractiveness, while all the men in that caste are at the 30th percentile in terms of, say, income.

[4] We estimate that its circulation represents about one sixth of the literate Bengali-speaking population of greater Kolkata.

equilibrium matching pattern we observe. That being said, this is a population that has the reputation to be more liberal than average; the preference for respecting caste rules we see within this group is thus probably a lower bound for what we would find in the general population.[5]

Our data collection from this group started with an interview, where we collected information on the prospective grooms and brides, about the letters they had received in response to their ad, their subjective ranking of those responses, and which of these responses they were planning to pursue further. We also asked them which ad in the newspaper they were planning to respond to themselves. At a second interview, a year later, we asked them whether they were married or engaged, and, if so, what were the characteristics of their (prospective or actual) spouses.

An advantage of this data set is thus that the entire information available to the ad-placer is also observed by us. At the time we initially interviewed them, ad-placers had just received the letter, and they had not yet met the prospective groom or bride or their parents.

The number of responses received in response to their ad, the set of ads they were planning to respond to, and the ranking and responses they assigned to the letters they had received, provide three independent ways to assess the relative importance given to different economic and social attributes (e.g., caste, education, beauty, proxies for wealth). For example, using either a linear probability or a fixed effect logit model, we estimate how the probability that an ad-placer decides to give further consideration to a response he received depends on the attributes of the ad-placer. the responder, and the interaction of the two.[6]

These alternative estimations methods lead to very similar qualitative conclusions: education, beauty, light skin, and high incomes are preferred. But the most striking result is that the preference for mating within one's own caste is strong: for example, we find in one specification that parents of a prospective bride would be willing to trade off the difference between no education and a master's degree to avoid marrying outside their caste. For men seeking brides, it is twice the effect of the difference between a self-described "very beautiful" woman and a self-described "decent-looking" one.

These results suggest that caste continue to play an extremely important role in structuring people's preferences for marriage partners in contemporary India, even among this educated, relatively affluent group. However, our estimates also clearly suggest that the caste preference is strongly "horizontal." We also find limited evidence of a preference for "marrying up" in terms of caste, in particular for women.[7]

[5] In the sample, 7 percent of the parents are from different castes, but 30 percent of their siblings married someone from a different caste.

[6] Our estimation strategy differs from that employed by Park (2007) and Fox (2007) who estimate preferences from equilibrium outcomes. Because of our rich data set, we are able to separately measure preferences and equilibrium outcomes rather than estimating the former by using the latter.

[7] These results are very similar to those obtained by randomly altering the income and caste of individuals in a series of matrimonial ads as in Dugar et al. (2008).

Our theoretical model would thus suggest that despite these strong caste preferences, the equilibrium matching patterns might not be strongly affected as long as the distribution of characteristics across castes is balanced. To explore this issue further, we observe both the actual matches from our survey data and also compute a set of stable matches that would be predicted to arise from the estimated preference parameters and the distribution observed in the set of ads collected. To compute such a set, we use a Gale-Shapley (Gale and Shapley 1962) algorithm. We estimate both which women remain single (in our sample men are on the short-side of the market) and who forms a union with whom. Hitsch et al. (2006) perform the same exercise for the online dating market in the United States while Lee (2007) does so in the context of a Korean match-making agency.

The point of this exercise is to check that we are doing a reasonable job of describing how the marriage market actually functions based on our inferred preferences. Encouragingly the set of stable matches approximates fairly well the set of actual marriages we observe in the data, with some exceptions, which we discuss later in the paper.

Reassured by this, we then perform several counterfactual exercises to examine how they affect the matching pattern generated by the Gale-Shapley algorithm. First, we compute the set of stable matches that would arise in our population if preferences were exactly as estimated above except that all caste variables were ignored. Our results indicate that the percentage of intra-caste marriages drops dramatically, implying that caste is not just a proxy for other characteristics households also care about–there are several potential matches for each individuals, both inside and outside his or her caste. At the same time, we also find that individuals are matched with spouses who are very similar on all non-caste characteristics to the partner they would have selected when caste was included within one's preferences. This suggest that caste has limited impact on matching patterns in equilibrium.

Second, we estimate the "equilibrium price" of caste in terms of a variety of attributes, defined as the difference between the spouses of two observationally identical individuals, one who is from the same caste and the other who is not. This is done by regressing a spousal characteristic, such as education, on all observable characteristics of the individuals and a dummy for whether the match is "within caste" among the set of simulated matches. There is no characteristic for which this measure of price is significantly positive.[8]

Thus, while individuals seem willing to pay large amounts in terms of attributes such as education and beauty to marry within their caste, they do not have to do so in equilibrium, because the caste preferences are horizontal. This implies that caste is not a significant constraint on the institution of marriage with respect to its role in matching individuals. Moreover this explains why the role of caste in marriage has not been weakened by economic forces: there is

[8] We demonstrate the power of this method by estimating positive "price" for other attributes, e.g. beauty.

not much of a trade-off between economic well-being and caste, so the aristocratic equilibrium can be persistent.

The remainder of the paper proceeds as follows: Section 2 first sketches a model where caste and other attributes interact on the marriage market. Section 3 presents the data while Section 4 elaborates on the methodology and the results of preference estimation. Section 5 highlights the results of the stable matches and Section 6 uses these results to derive conclusions regarding the equilibrium. Finally, Section 7 concludes.

2 Model

In this section we develop a simple model of marriage. Our goal is to identify some useful properties of the choice problem faced by decision-makers in the marriage market as well as the equilibrium matching pattern, in a world where people care about the caste of their partners, as well as some standard characteristics (e.g., education, beauty). These will motivate our empirical analysis and help us interpret some of the results.

2.1 Set up

Men and women are differentiated by "caste." The caste of an individual is $i \in \{1, 2\}$. They are ranked in descending order: $i = 1$ is the highest caste, followed by $i = 2$.

Men and women are also differentiated according to a "vertical" characteristic that affects their attractiveness to a potential partner. The characteristic of men will be denoted by $x \in [H, L]$ and the characteristic of women will be denoted by $y \in [H, L]$. We can think of these as education levels of men and women, or, income and beauty.

The payoffs of men and women are both governed by the quality of the match. We assume that this has two (multiplicatively) separable elements, one governed by the vertical characteristics, $f(x, y)$, and the other by caste, $A(i, j)$.

We assume that the function $f(x, y)$ is increasing with respect to both arguments and supermodular. Thus, other things constant, everyone prefers a higher attribute partner. Also, following the tradition of Becker, these characteristics are assumed to be complementary in the payoff of men and women.

The function $A(i, j)$ captures the quality of a match for a individual of caste i (man or woman) who is matched with a partner of type j. This is defined as follows:

$$A(i, j) = 1 + \alpha\{\beta(2 - j) - \gamma(i - j)^2\}$$

where $\alpha \geq 0, \beta \geq 0, \gamma \geq 0$. It is readily verified that $A(i, j) \geq 0$ as long as $\alpha\gamma < 1$ (which we

assume) and as long as $\gamma > 0$ the function displays strict complementarity with respect to caste: $\frac{\partial^2 A(i,j)}{\partial i \partial j} > 0$.

This caste-based match quality function is flexible. It allows there being a vertical as well as a horizontal component to caste. For example, if $\beta = 0$ then caste is purely horizontal: people want to match within their caste. Otherwise, the higher the caste of the partner (lower is j) the higher the match specific gain to an individual of caste i. On the other hand, if $\gamma = 0$ then caste is purely vertical with everyone preferring a higher caste partner, as in Anderson (2003). In the marriage literature, a high β will be viewed as leading to hypergamy and a high γ will be viewed as leading to endogamy.

We also assume that some members of the population, drawing from both caste groups, have caste-neutral preferences, $\alpha = 0$. These individuals put no weight on the caste of a potential partner, i.e., for them $A(i,j) = 1$ for all $i = 1,2$ and $j = 1,2$. For those who are caste-conscious, they value a caste-neutral individual of caste i ($i = 1,2$) in the same way as they would a caste-conscious individual of caste i ($i = 1,2$).

Given these two elements governing the quality of a match, we assume that the payoff of an individual of gender G, of caste i who is matched with someone of caste j in a union where the man's quality is given by x and that of the woman's by y is given

$$u^G(i,j,x,y) = A(i,j)f(x,y) \text{ for } G = M,W$$

Several observations are in order. First, we assume that the non-caste component of the quality of a match, $f(x,y)$, is the *same* for a man and a woman. This is clearly most relevant to settings where this aspect of a match is a pure public good (e.g, children, joint activities).[9]

Second, the caste and the non-caste components interact with each other: in particular, a "good" caste-specific match will have higher marginal product of the non-caste attributes.

Third, the caste matching function is symmetric for men and women. That is, a man of caste 1 marrying a woman of caste 2 gets the same payoff that a woman of caste 1 would get from marrying a man of caste 2.

2.2 The role of dowry: adding unobserved characteristics

A key modeling decision is whether to assume that we are in a non-transferable utility (NTU) environment (as in studies of the United States matching market studied for example by Hitsch et al. 2006, Fisman et al. 2006 and Fisman et al. 2008) or the transferrable utility (TU) environment, which is more traditional in the literature (e.g., Becker 1973, Lam 1988).

[9]In a NTU world, if men and women get very different payoffs from the standard component of a match, it is hard to provide much in the way of characterization. In any case, our results go through if men and women put different weights on the standard component of a match but these weights are not very different.

This may appear surprising in India, a setting where there are dowries. However, dowries are frowned upon given the prevailing social norms in this group, i.e., middle-class, educated, urban Bengalis. They are also illegal, which made it impossible for us to collect data on them: no one is observed in our data set asking for a dowry or offering one. While it is hard to deny that the practice exists, the standard view is that dowries are not particularly important in this population.[10] This inclines us toward the NTU approach.

However, the TU environment can be relevant even in the absence of dowries or brideprice, so long as there is some other "currency" which can be used to make ex ante transfers (e.g., the decision about where the couple should live, how much should be spent on the wedding, etc.). Furthermore dowries or other financial transfers may still exist even if they are not obseved. To not entirely foreclose the possibility of transfers, we take the following approach: our estimation of preferences is based on recording the observable characteristics of those who get chosen (to get a call back or a letter) out of a set of "applicants." We first observe that as long as there are enough people who prefer not to demand transfers (a not insignificant part of our sample actually spend money in the form of ad space to explicitly mention that they do *not* want a dowry), it makes sense to first choose everyone whom you would have chosen *ignoring the possibility of their asking for a dowry or offering one*, and to actually find out whether or not they want a dowry (or want to offer one) by contacting them. They can then discard the ones who ask for too much or offer too little based on better information. Obviously this logic only works if the cost of contacting another person is small which, given the large numbers people contact, seems plausible.

Proposition 4 in the appendix formally states this argument for the case where there is one unobservable variable (e.g., a dowry demand/offer) that is potentially correlated with the observables.[11]

We therefore only model the NTU world. Assuming that the conditions detailed in the appendix hold (namely, the exploration costs are not too high), what we observe in the data is people's true ordering between those whom they consider and those whom they reject. Based

[10]We have so far failed to locate a study on dowry in this population that would throw light on its extent. However, we note that while Kolkata has 12 percent of the population of the largest metropolitan cities in India, it has only 1.9 percent of the so-called "dowry deaths" in these cities (about 6,000 in a year, India-wide), which are episodes where a bride is killed or driven to commit suicide by her in-laws following negotiation failure about the dowry. To the extent that the prevalence of dowry death partly reflects the prevalence of dowry, it suggests that they are less prevalent in Kolkata than in other major cities in India.

[11]The assumption in this Proposition is that the unobserved attribute has a fixed value. In that, it is more like charm than like a demand for dowry, which is something that might adjusts to exactly compensate for differences in other attributes. Nevertheless as long as each set of candidates with the same observable characteristics contains a sufficiently large subset which is on average identical to the rest of the group in everything except for the fact it will not accept a dowry, and as long as it is not possible to predict this in advance (dowry demands or offer are not made in writing), it makes sense to rank everyone as if no one wants a dowry, as long as the cost of search is not too large.

on this ranking we infer people's preferences over a range of attributes. We will come back to discuss some direct tests of the no-dowry assumption.

2.3 The price of caste

In the data we observe the trade-offs people make between caste and other observables in selecting the set of people they are prepared to explore further. Here we want to develop a simple notion of the "price" of caste that corresponds to this trade-off, i.e., the extent of partner quality one is willing to give up to marry within caste. Consider a man of type x who belongs to caste 1. Suppose the best match he has is a woman of quality y from his own caste. Then he is indifferent between marrying a woman of quality y within his own caste and a woman of caste 2 if the attribute of this woman is higher by the margin ε given by:

$$(1+\alpha\beta)f(x,y) = (1-\alpha\gamma)f(x,y+\varepsilon). \tag{1}$$

We can solve $\varepsilon(x,y,\beta,\gamma)$ from this equation. This can be interpreted as the "supply" price of caste: this is the price at which a high caste person (here, a man) will agree to marry a low caste person. Clearly the supply price is zero when $\alpha = 0$.

Similarly consider a woman of type y' who belongs to caste 2. Suppose the best match she can find in her own caste group is x'. Then she is indifferent between marrying a man of quality x' within her own caste and a man of caste 1 if the attribute of this man is not lower than the margin δ:

$$(1+\alpha\beta-\alpha\gamma)f(x'-\delta,y') = f(x',y').$$

We can solve $\delta(x',y',\beta,\gamma)$ from this equation. This can be interpreted as the "demand" price of caste: this is the price a person of low caste is willing to pay to marry a higher caste person. As before, for $\alpha = 0$, the demand price of caste is 0.

The two following observations follow immediately from the fact that f is increasing in both arguments and the definition of the supply and the demand prices that:

Observation 1 *If $\beta = 0$ (a purely horizontal world), $\delta \leq 0 \leq \varepsilon$, whereas if $\gamma = 0$ (a purely vertical world), $\delta \geq 0, \varepsilon \geq 0$ for all $\beta > 0$.*

Observation 2 *The supply price of caste is increasing in β and γ, whereas the demand price of caste is increasing in β and decreasing in γ.*

Observing a high supply price is consistent with both strongly vertical and strongly horizontal preferences. By contrast a high demand price suggests that preferences are vertical. This feature of the model will be important in interpreting our empirical results.

Once we have the concepts of demand price and supply price, the following implication is straightforward:

Observation 3 *An inter-caste marriage takes place if and only if* $\varepsilon \leq \delta$.

That is, the quality gain a man (woman) needs to marry down cannot exceed the quality loss a woman (man) is willing to tolerate for marrying up.

Together these three observations suggest that inter-caste marriages are more likely in a world where caste is more vertical. We now turn to this in more detail.

2.4 Matching in a balanced population

Other than preferences, the distribution of the population in terms of caste and quality would clearly affect the equilibrium matching pattern and the associated equilibrium price of caste. We begin our analysis by focusing only on the role of preferences.

Let the distribution of x and y within each caste be balanced. In other words, if m_k^i is the number of men of type k $(k = L, H)$ in caste i and w_k^i is the number of women of type k $(k = L, H)$ in caste i then $m_k^i = w_k^i$ for all $k = L, H$ and for all $i \in \{1, 2\}$. More intuitively, the assumption implies that any man whose type is z $(z = L, H)$ in caste i can find a woman whose caste is i and whose type is z.[12]

We begin with the following simple observation:

Observation 4 *With balanced population within each caste group, if marriage is restricted to within caste, the equilibrium displays assortative matching.*

Since the thought experiment is to restrict attention to within-caste matches only, this result follows immediately from the assumption of $f(x, y)$ being increasing in both arguments. If an L-type man is matched with an H-type woman (or vice versa) somewhere else an H-type man must be matched with an L-type woman, and this assignment cannot be stable as an H-type woman and an H-type man can form a pair that will make them both better off.[13]

Let us consider the possibility of inter-caste marriage. With a balanced population it is always possible to find a match within your own caste if that is what you want. Among caste-conscious individuals, the only possible inter-caste marriage will occur between a H-type person

[12] It is worth emphasizing here that nothing rides on the male and female characteristics being both labeled L, H. The male characteristic could be height and the female characteristics could be education; our assumption is that there are as many tall men as there are well educated women. It remains that this is a strong assumption. We will come back briefly to what would happen if it fails.

[13] This is under the assumption of NTU. With TU, as is well known from Becker (1973), to get positive assortative matching, x and y would need to be strict complements.

from caste 2 and a L-type person from caste 1. No other caste-conscious individuals would find it profitable to search for a different partner.

Males of H-type from caste 2 will find it profitable to marry outside their caste if

$$f(H,H) \leq A(2,1) f(H,L).$$

A female of H-type from caste 2 will find it profitable to marry outside caste if

$$f(H,H) \leq A(2,1) f(L,H).$$

As long as $\alpha > 0$, these conditions can be rewritten as:

$$\gamma \leq \beta - \frac{1}{\alpha}\left(\frac{f(H,H)}{f(H,L)} - 1\right) \tag{2}$$

$$\gamma \leq \beta - \frac{1}{\alpha}\left(\frac{f(H,H)}{f(L,H)} - 1\right). \tag{3}$$

When these conditions are not satisfied, no caste-conscious individuals will match outside their caste. Notice that $\frac{f(H,H)}{f(H,L)} > 1$ and $\frac{f(H,H)}{f(L,H)} > 1$. Thus, when $\beta \leq \gamma$, no one will marry outside his or her caste and we will observe assortative matching in equilibrium, which is also what we would observe if caste was entirely irrelevant:

Proposition 1 *With balanced population within each caste group, if the horizontal component in preferences, γ, is at least as important as the vertical component β, i.e., $\gamma \geq \beta$: (i) inter-caste marriages can never take place that involve at least one caste-conscious individual ($\alpha > 0$); (ii) those with caste-neutral preferences are indifferent between marrying within caste or outside; (iii) the equilibrium displays assortative matching and so the equilibrium price of caste is zero.*

Proof. (i) Already shown

(ii) This follows directly from the balanced population assumption and the fact that $\alpha = 0$.

(iii) Given (i) and (ii) there is no strict incentive to marry outside caste (caste-neutral individuals may be indifferent) and given the balanced population assumption within each caste group, assortative matching results. This immediately implies that the equilibrium price of caste is zero: we would not observe an individual sacrificing partner quality in order to marry outside caste. ■

With sufficiently horizontal preferences ($\gamma \geq \beta$) and a balanced population, the only inter-caste marriages are between those who do not care about caste.[14]

[14]Since individuals are indifferent, other idiosyncratic factors can play a tie-breaking role and lead to inter-caste marriages.

We now turn to the case where inter-caste marriages may emerge in equilibrium even with balanced populations. From the above results we know that for this to happen, it must be the case where β is relatively large compared to γ (i.e., caste preferences are primarily vertical, not horizontal).

In this case, two types of equilibrium can arise. If conditions (2) and (3) hold, then H-type men and women from caste 2 will be willing to marry L-type women and men from caste 1. Caste-neutral L-type individuals from caste 1 will be delighted to enter in this pairing. However, caste-conscious L-type individuals from caste 1 will want to enter this union as long as:

$$A(1,1)f(L,L) \leq A(1,2)f(H,L)$$

for females and for males, if

$$A\left(1,1\right)f\left(L,L\right) \leq A(1,2)f(L,H).$$

Replacing the caste-specific payoffs with their actual values allows us to rewrite these conditions as:

$$\gamma \leq \frac{1}{\alpha}\left(1-\frac{f\left(L,L\right)}{f\left(H,L\right)}\right)-\beta\frac{f\left(L,L\right)}{f\left(H,L\right)} \tag{4}$$

$$\gamma \leq \frac{1}{\alpha}\left(1-\frac{f\left(L,L\right)}{f\left(L,H\right)}\right)-\beta\frac{f\left(L,L\right)}{f\left(L,H\right)}. \tag{5}$$

As $1+\alpha\beta-\alpha\gamma<\frac{1+\alpha\beta}{1-\alpha\gamma}$, (2) and (4) can be combined as

$$\frac{f(H,H)}{f(H,L)} \leq 1+\alpha\beta-\alpha\gamma<\frac{1+\alpha\beta}{1-\alpha\gamma} \leq \frac{f(H,L)}{f(L,L)} \tag{6}$$

and, analogously, (3) and (5) can be combined as

$$\frac{f(H,H)}{f(L,H)} \leq 1+\alpha\beta-\alpha\gamma<\frac{1+\alpha\beta}{1-\alpha\gamma} \leq \frac{f(L,H)}{f(L,L)}. \tag{7}$$

A necessary (but not sufficient) condition for these conditions to hold is that $f(H,H)f(L,L)<\min[\{f\left(H,L\right)\}^2,f\left(L,H\right)\}^2]$. This is satisfied, for example, by a CES function $f\left(x,y\right)=\left(x^a+y^a\right)^{\frac{1}{a}}$ when $a>0$.[15]

We are now ready to formally characterize the observed matchings under a balanced population assumption and more vertical caste preferences:

[15] In this case, the condition is given by $(2^{1/a}H)(2^{1/a}L) \leq (H^a+L^a)^{\frac{2}{a}}$. This holds because $4H^aL^a \leq (H^a+L^a)^2$ and thus $4H^aL^a \leq (H^2a+L^2a+2H^aL^a)$. This can be rewritten as $0 \leq (H^2a+L^2a-2H^aL^a)=(H^a-L^a)^2$. This will hold for any $a>0$.

Proposition 2 *Assuming the size of the caste-neutral group being small: (i) Inter-caste marriages involving a caste-conscious individual of caste 2 and a caste-neutral individual of caste 1 who is of lower quality will take place when γ is small relative to β.*

(ii) Inter-caste marriages involving a caste-conscious individual of caste 2 and a caste-conscious individual of caste 1 who is of lower quality will take place when γ is small, and β is not too large.

(iii) The equilibrium price of caste will be positive and will decrease the greater the share of caste-neutral individuals.

Proof. (i) The relevant condition is (2) for men and (3) for women. It will be satisfied if β is large enough relative to γ.

(ii) The relevant condition is (6) for a caste-conscious man of H-type in caste 2 and a caste-conscious woman of L-type in caste 1 to get married Switching the gender, the relevant condition is (7). It is clear that for very high values of β the latter set of conditions cannot hold as $\gamma \geq 0$. Subject to β being not too low, there is a threshold level of γ such that if γ is smaller than this value, then both conditions will hold.

(iii) Since there will be non-assortative matching under the conditions stipulated in (i) and (ii), the equilibrium price of caste will be positive: some high quality individuals of caste 2 will marry low quality individuals of caste 1. Since we have two quality levels, the effect of the size of the caste neutral population on the equilibrium price of caste is discrete: as it goes up above a certain threshold, all caste 2 individuals who want to marry up in caste will find a caste-neutral caste 1 individual of the same quality and so the price of caste will be zero. Otherwise it will be positive. ■

Furthermore, one can make the following observation about the matching patterns:

Observation 5 *Caste-conscious, high quality men and women in the upper caste and low quality men and women in the lower caste will marry within caste and assortatively. Observed inter-caste marriages among caste-conscious individuals will take place between low quality men (women) of the high caste and high quality women (men) of the lower caste.*

2.5 Matching in an unbalanced population

The simple vertical-horizontal dichotomy of the previous section is only possible because we assumed a balanced population. In general, the distribution of the population will affect the equilibrium outcomes. Without balanced population, even a high caste person who is of the high type may not find a corresponding high type person with her own caste, and therefore has to choose between a low type person of her own caste and a high type person of the lower caste. Therefore she might marry out of caste even if her preferences are entirely horizontal ($\beta = 0$).

To highlight this effect we focus on the case where preferences are purely horizontal (i.e., $\gamma > \beta = 0$) so that in a balanced population matches will be assortative, and no inter-caste marriages will take place. Also, let us assume that everyone is caste-conscious ($\alpha > 0$). Then, H-type males who find H-type females from within the same caste marry within caste.

For those who don't find an H type within caste 1, their choices are to marry an L-type female from within the same caste or an H-type female from caste 2. The latter is more attractive if:

$$(1 - \alpha\gamma)f(H,H) \geq f(H,L)$$

or

$$\frac{1 - \frac{f(H,L)}{f(H,H)}}{\alpha} = \bar{\gamma} \geq \gamma.$$

There will be a similarly defined cut-off value for γ at which an H-type female from caste 2 to agree to marry this individual. Assuming the payoff from being single to be zero, for an L-type individual in caste i who cannot find an L-type individual of the opposite sex within the same caste (and, by transitivity, an H-type person of the opposite sex within the same caste) will be willing to marry an L-type individual of the opposite sex from caste $j \neq i$. The latter will agree if he/she too cannot find an L-type match from his or her own caste group. The payoff of both parties will be $(1 - \alpha\gamma)f(L,L) > 0$ (as we assume $\alpha\gamma < 1$).

Recall that a balanced population assumption implies that $m_k^i = w_k^i$ for all $k = L, H$ and for all $i \in \{1, 2\}$. If $m_k^i > w_k^i$ and $w_k^j > m_k^j$ for some k ($k = L, H$) and $i \neq j$ then we define the sex ratio for quality level k to be complementary across the two caste groups. Now we are ready to state:

Proposition 3 *With an unbalanced population, and complementary inter-caste sex ratios for at least some quality level k, inter-caste marriages will take place even with purely horizontal preferences ($\gamma > 0 = \beta$) if $\gamma \leq \bar{\gamma}$. Inter-caste marriages, if they take place, will be assortative and the equilibrium price of caste will be zero.*

Proof. This follows from the fact that given the assumption $\gamma \leq \bar{\gamma}$, an H-type man in caste i prefers to marry an H-type woman in caste j rather than marrying an L-type woman in caste i, and vice versa. Also, as $\bar{\gamma} < \frac{1}{\alpha}$, an L-type man in caste i and an L-type woman in caste j prefer marrying each other rather than staying single. Given this assortative matching directly follows, and so the equilibrium price of caste will be zero. ■

Therefore in the unbalanced population case, so long as sex ratios are complementary across caste groups for at least some quality level, there will be inter-caste marriages even with purely horizontal preferences. $\beta > 0$ will reinforce this tendency. If sex ratios are not complementary

for any quality level then not a lot can be said in general. Among other factors, the outcome would depend on the aggregate sex ratio.

2.6 Discussion

First, with horizontal preferences and with a balanced population ($\beta < \gamma$), there are no inter-caste marriages between people who care about caste ($\alpha > 0$). Moreover, in this case, if everyone became caste neutral (i.e., $\alpha = 0$ so that for all i and j, $A(i,j) = 1$) the same pattern of matching will be observed in terms of all the attributes except caste. If we observe inter-caste marriages they must take place between caste-neutral people, and the equilibrium price of caste is zero.

Second, contrast this with a world where preferences are significantly vertical (i.e., as in Proposition 2). Now inter-caste marriages will take place. In this case, if everyone becomes caste-neutral, there will be significant changes in the pattern of matching as now there will be assortative matching in terms of x and y for the whole population. In that case, the equilibrium price of caste will be positive – people will be willing to "pay" in terms of partner quality to marry up in terms of caste.

Third, when the population is not balanced, inter-caste marriages can occur even with purely horizontal preferences. In this case, if preferences are sufficiently horizontal, inter-caste marriages will tend to be assortative and the equilibrium price of caste will tend to be low.

Given these theoretical predictions, the empirical sections that follow will focus on estimating the magnitude of the caste preferences in our sample and determining whether they are horizontal or vertical. Then, using these estimates, we will explore the equilibrium consequences that these caste preferences generate for marital pairing.

3 Setting and data

3.1 Setting: the search process

Our starting point is the set of all matrimonial ads placed in the Sunday edition of the main Bengali newspaper, the *Anandabazar Patrika* (ABP), from October 2002 to March 2003. With a circulation of 1.2 million, ABP is the largest single edition newspaper in India and it runs a popular special matrimonial section every Sunday. The search process works as follows.

First, the parents or relatives of a prospective bride or groom place an ad in the newspaper. Each ad indicates a PO box (provided by the newspaper), and sometimes a phone number, for interested parties to reply. They then get responses over the next few months (by mail or by phone), and elect whether or not to follow up with a particular response. While ads are placed

by both sides of the market, "groom wanted" ads represent almost 63 percent of all ads placed.

When both parties are interested, the set of parents meet, then the prospective brides and grooms meet. The process takes time: in our sample, within a year of placing an ad, 44 percent of our sample of ad-placers whom we interviewed, were married or engaged although most had placed only a single ad. Of those who got married, 65 percent met through an ad, the rest met through relatives or, in 20 percent of the cases, on their own (which are referred to as "love marriages ").

3.2 Sample and data collection

We first coded the information in all the ads published in the Sunday edition over this time period. We excluded ads placed under the heading "Christian" or "Muslims" in the newspaper given our focus on caste, which is primarily (though not exclusively) a phenomenon among Hindus. The details on the information provided and the way it was coded are provided below. We refer to this data set of 22,210 ads as the "ad-placer sample."

We further restricted our attention to ads that did not mention a phone number, and requested all responses to be sent at the newspaper PO Box or to a personal mailing address.[16] This restriction was necessary to make sure that the letters received in response to an ad reflect all the relevant information the ad-placer has on the respondent. About 43 percent of the ad-placer sample included a phone number (sometimes in addition to a PO Box and sometimes as the only way to contact the ad-placer). We find little differences between the characteristics of the ads that included a phone number and those that did not, except in terms of geographical location: fewer ad placers with phone numbers were from Kolkata.

After excluding these ads from the ad-placer sample, we randomly sampled 784 ads. With ABP's authorization, respondents were approached and asked whether they would agree to be interviewed when they came to collect the answers to their ads at the newspaper PO Box. Only one sampled respondent refused to be interviewed. The ads placed by the 783 individuals who completed the survey form the "interview sample."

The interview was conducted in the ad-placer's home after a few days with the person in charge of the search, usually the parent, uncle or older brother of the prospective groom or bride. Detailed information was collected on the prospective groom or bride, his family and the search process for a marriage partner.[17] In particular, ad-placers were asked whether they also replied to other ads and, when they did, to identify the ads they had responded to among the ads published in the past few weeks. Ad placers were also asked how many letters they received

[16]Only a small fraction of ads included only a personal mailing address (namely, 4 percent of our interview-sample, and 8 percent of the ad placer sample).

[17]The questionnaire is available online at http://www.econ.umd.edu/ Lafortune/Questionnaire/.

in response to their ad (on average 83 for bride-wanted and 23 for groom-wanted ad placers), and to identify the letters they were planning to follow up with (the "considered" letters). We then randomly sampled five letters from the set of "considered" letters (or took the entire set if they had less than five in this category), and ten (or all of them if they had less than ten in this category) from the set of the "non-considered" letters, and requested authorization to photocopy them. The information in these letters was subsequently coded, using the procedure outlined below. We refer to this data set as the "letter data set."

Finally, a year after the first visit, this original interview-sample was re-interviewed, and we collected information regarding their current marital status and their partner's choice. Only 33 ad-placers out of the entire sample could not be contacted. Out of those we reached, 346 were married or engaged, and 289 of those agreed to a follow-up interview and gave us detailed information regarding their selected spouse, the date of the marriage and their overall search process including the number of ads posted and the way the match was made. Appendix Tables C.1 and C.2 compares ad-placers found and not found and those who agreed or refused to answer the follow up questions. There appears to be little systematic differences between the two groups.

3.3 Variable construction

Ads and letters provide very rich and mostly qualitative information. A data appendix describes the coding process. In this subsection, we only discuss the coding process for the caste information.

If caste was explicitly mentioned in the ad or letter, we used that information as the caste of the person. Caste is often not explicitly mentioned in the ad because the ad is usually placed underneath a particular heading in the newspaper corresponding to a caste (or sometimes a small group). If caste is not directly mentioned in the ad, the heading is used for this classification. The information on caste is readily available, directly or indirectly, in the overwhelming majority of ads (98 percent). In the letters, caste is explicitly mentioned in about 70 percent of the cases.

There are numerous castes and sub-castes in India. Ad-placers or letters can be more or less specific in identifying themselves. Historically there was a more or less clear hierarchy among the broad caste groups, but within each broad group, there was no clear ordering. We therefore grouped castes into eight ordered broad-caste groups, based on the classifications in Risley (1981) and Bose (1958), with Brahmin at the top (with the rank of 8, and various schedule castes at the bottom, with the rank of 1). Appendix Table C.3 presents the classification.

To determine whether a letter writer and an ad-placer are from the same caste, we attributed to each letter or ad the specific sub-caste mentioned in the ad. If the ad-placer or letter writer only mentioned a broad group, he or she is assumed to be from any of the specific sub-castes.

For example, a self-identified Kulin Brahmin is considered to be from a different caste as a self-identified Nath Brahmin (though the vertical distance between them is set to zero), but is considered to be of the same caste as someone who simply identified himself as a Brahmin.

Another relevant piece of information is the stated preferences regarding caste. Among the sampled ads, more than 30 percent of individuals specify their preference for marrying within their caste (using phrases such as "Brahmin bride wanted"). Another 20 to 30 percent explicitly specify their willingness to unions outside their own caste by the use of phrases such as "caste no bar." The remaining 40 to 50 percent do not make any mention of preferences regarding caste.

The remaining variables coded were: education (in 7 categories), earnings and occupation for men (we construct an occupational score, referred to as "wage" in what follows), family origin, physical characteristics, and some more rarely mentioned traits (astrological signs, blood types, etc.). The data appendix provides more details on the coding.

3.4 Summary statistics

Table 1 presents summary statistics for both our interview sample and the full set of ads. The two samples look quite similar, except that the interview sample is more likely to live in Kolkata (the Kolkata sample was less likely to provide a phone number).

Our sample is drawn mostly from the Bengali middle class, as evidenced both by the prevalence of higher caste individuals (a quarter of the sample are Brahmin), and educational achievement. Education levels are mentioned in the ad by 90 percent of women and 80 percent of men. Almost all men and women (90 percent) have at least a bachelor's degree. Women rarely mention their occupation. When they do, their occupational score (5.54 for the ad-placer sample and 5.55 among the interview sample) is similar to that of men (5.20 for the ad-placer sample and 5.60 for the interview sample) and significantly higher than the median urban formal sector occupational score (from Bargain et al. 2007 and Glinskaya and Lokshin 2005). This group enters the marriage market after they have completed their education and (at least for men) found a job: the average age is 27 for women, and 32 for men. Around 50 percent of the sample lives or works in Kolkata and slightly less than half consider their family as originating from West Bengal.

Physical characteristics clearly play an important role in the marriage market. Height is mentioned in the ad by 96 percent of the women and 90 percent of the men. Skin tone is mentioned in 75 percent of the cases, beauty, in over 70 percent of the ads. There does not appear to be much boasting about physical appearance, however. More ads describe the bride as being "decent-looking" than either "beautiful" or "very beautiful."

Table 2 presents similar statistics for two different samples: the sample of people who wrote a letter in response to an ad ("the letter writers") and the sample of actual spouses, collected

from interviews with the ad-placer.[18] A few prospective grooms (7 percent) explicitly mention that they will not demand a dowry. No one mentions that they want a dowry.

This table also shows comparisons between the ad-placers and the letters they have received, as well as with their eventual spouses. In this table, as well as in the remainder of the paper, all differences are presented in terms of the difference between the characteristics of the man and the characteristics of the woman.[19]

Two-thirds of the letters that mention caste are from someone from the same caste as the ad-placer. The fraction of within-caste marriages among actual matches is a little higher than the fraction of letters that come from within one's caste: 72 percent of the prospective grooms and 68 percent of the prospective brides who are married after a year have married within their own narrow caste. This fraction increases to 76 percent and 72 percent respectively if we use the broad classification in terms of caste. Men who marry outside of caste tend to marry women from a lower caste while women who marry outside of caste tend to marry someone from a higher caste. Women tend to marry grooms who have either the same education (42 percent) or who are more educated than them (45 percent). Men are more likely to marry similarly or more educated women than themselves and 72 percent to 75 percent of the brides and grooms are from the same family origin (i.e., West or East Bengal).

4 Estimating preferences

Using this data, we now estimate the preferences over various characteristics, exploiting the choices made by ad-placers and people who replied to their ads. We first discuss our basic empirical strategy and present the results. We then empirically examine various concerns about why the coefficients we observe may not actually represent households' preferences.

4.1 Basic empirical strategy

The first goal of this section is to estimate relative preferences for various attributes in a prospective spouse.

We assume that the value of a spouse j to a particular individual i can be described by the following function:

$$U(X_j, X_i) = \alpha X_j + \beta f(X_i, X_j) + \mu_i + \varepsilon_{ij} \tag{8}$$

where α captures the effect of the characteristics of person j, β specifies how this effect might be different depending on person's i own characteristics and μ_i represents ad-placer fixed effects.

[18] Few families could show us the original ad or letter of the spouse.

[19] Since the sampling was stratified with unequal weights, each letter is weighted by the inverse of its probability of selection.

We have in our data several indications of individuals' revealed preference for one potential spouse over another that can allow us to estimate the parameters of equation (8).

First, we know whether an ad-placer is following up with a particular letter writer or not. We thus have information that he preferred this letter to the letters he did not consider. Second, the ad-placers also provided us with their ranking of each letter we sampled. In addition, for ad-placers who have themselves replied to ads, we know which ads they decided to reply to (and we also know the universe of ads they could have replied to on that particular date). Furthermore, we know that a letter writer decided to reply to an ad. Finally, we also know how many responses an ad received.

We focus on the decision of the ad-placer to respond to a particular letter. First, as we will explain later, the results using the ranking or the binary decision of replying to a letter are extremely similar. Second, whether an ad-placer responds to a letter has two advantages over the three last options. First, we can be sure that the ad-placers have read all the letters they have received, so the set over which choices are made is well-defined. Second, strategic behavior is *a priori* less likely in this sample since the letter writer has already expressed interest in the ad-placer. The results from these other strategies will be presented in the appendix. The results are very consistent, but we will highlight the main differences below.

The regressions we estimate thus take takes the following form:

$$y_{ij} = \alpha X_j + \beta f\left(X_i, X_j\right) + \upsilon_i + \epsilon_{ij}, \tag{9}$$

where y_{ij} is a dummy equal to 1 if ad-placer i replied to letter j.[20] In the empirical analysis, we specify $f(X_i, X_j)$ to include dummies for whether the value of some elements of the X vector are equal for i and j (for education, caste, location), the difference between the value of the elements of the vector for some attributes (always normalized such that we take out the average difference between men and women), and its square.[21] We estimate equation (9) using a conditional logit with fixed-effects for each person i, and OLS with fixed effects.[22] Note that for characteristics of ad-placers, we could use either the information provided in their ad or their response to our interview questions. In order to use these estimates in the stable matching exercises that follow, the former was employed. However, very similar results were obtained when using the interview data.

[20] This is similar to the regression framework of Hitsch et al. (2006).

[21] For linear variables, such as age or height, we include only the difference between the value of the variable for the man and the woman and its square, not the level of age or height for the letter writer: this is because once we include a fixed effect for the ad-placer, the age of the letter writer and the difference in age are co-linear.

[22] The exact likelihood is not a logit because of the sampling procedure described above. However, it is reassuring that the results are roughly the same between the OLS and logit estimators.

4.2 Results

Table 3 presents the results of fixed-effects and conditional logit regressions, where the binary decision of whether or not an ad-placer i responded to a letter j is regressed on a set of characteristics of the letter, and its interactions with those of the ad.

Columns 1 to 5 present the specifications for groom-wanted ads, and columns 6 to 10 present the specifications for bride-wanted ads. Recall that in both cases, differences are presented in terms of the difference between the characteristics of the man and the characteristics of the woman. A positive difference in education for example, means that the prospective groom is more educated than the prospective bride. Also a positive difference between the man's and woman's caste indicates that the man is of a higher caste. A variable is set to zero if the letter did not mention that characteristic, and we include a dummy variable to indicate a missing characteristic.[23]

Most attributes have the expected signs in the utility function: both women and men prefer more educated spouses; science and commerce are the preferred fields. Women prefer men with higher incomes. Men prefer younger women. and women prefer men their own age. Both dislike large differences in age. As Hitsch et al. (2006), we find that looks matter: men prefer women who describe themselves as beautiful or very beautiful, and seem to have a strong preference for lighter-skinned brides. For example, the OLS estimate suggests that the probability to be called back would be higher for a very light-skinned woman without an education than for a dark-skinned woman with a college degree. Both men and women prefer a spouse who lives in Kolkata (recall that a majority of our families are from Kolkata), and with similar family origin (i.e., East or West Bengal).

Caste plays a very prominent role. In particular, both men and women seem to have a very strong preference for marrying within the same caste. The OLS estimates indicate that a woman is 13 percentage points more likely to call back a prospective groom if he is from the same caste, controlling for all other attributes. A man is 17 percentage points more likely to call back a woman from his caste. These are large differences, considering that the average call back rate is about 28 percent. These results also indicate a high preference for caste relative to other attributes. For example, in the bride-wanted ad the probability to be called back is the same for a man from the same caste and no education as that for a man from a different caste with a master's degree. Men are willing to sacrifice three shade of skin tones to marry someone within their caste (column 6). Comparing the trade-offs implied by the coefficients on caste and other characteristics in the OLS and logit specification, we find them to be very similar.

[23] All models were estimated with and without including a series of additional covariates (for example, how "cultured" the family is, its wealth level, astrological sign). To save space we focus on the more parsimonious specification in the tables; the results are extremely similar when these additional controls are included.

Given our theoretical framework, an important issue is whether preference for caste is horizontal or vertical. Conditional on marrying out of their caste, women prefer men who are as close to their caste as possible, while among men, those from relatively high castes want women from a very similar caste, while men from a relatively low caste prefer women from the highest available caste. The magnitudes of the coefficient on the difference in caste, however, are much smaller than those for being of the same caste.

One possibility is that several of the variables in these regressions are co-linear proxies for the same underlying attribute. Specifically, the basic specification includes income (when reported), education, type of degree, and occupational score (when reported). This may artificially depress the coefficient of these variables relative to the caste variable. To investigate this possibility, we estimate in columns (4) and (9) a more parsimonious specification. We first regressed the log income of the letter writer (when reported) on all the education variables and the occupational score (including dummies when not reported). We then constructed for each ad-placer and letter writer a "predicted income" measure using the coefficients of that regression, and included this variable instead of all the education, income, and wage variables adjusting the standard errors for the fact that this regressor is generated by using the method suggested by Murphy and Topel (1985). Predicted income has a strong and significant impact on the probability of call back, but this does not shrink the relative importance of caste. A woman from a given caste would be as likely to contact a male from her own caste with a given predicted income level than a male from a different caste who is predicted to earn 50 percent more.

To display graphically the trade-off between caste and other attributes, Figures 1 and 2 show indifference curves, drawn using the conditional logit estimates.[24] They display the age difference, height difference, education, and income a prospective spouse needs to have to keep the ad-placer indifferent when his or her caste changes, expressed in standard deviations.

In both cases, the demand price for caste is quite substantial. To remain indifferent between two prospective brides, one of the same caste and one from a caste one notch below, the second one must have 3 standard deviations more education, must be 5 standard deviations more closer in age or earn 6 standard deviations more income. The differences are slightly less marked for preferences of women but still very marked for same caste. For both genders, a smaller penalty is attached to marrying individuals of a higher caste than of a lower one, in addition to the penalty of marrying outside one's caste. This is related to the findings of Fisman et al. (2008) who find strong same-race preferences among female speed daters that is unrelated to physical attractiveness. Similarly, Hitsch et al. (2006) also find same-race preferences, particularly for women.

[24]The displayed graphs were generated for a Kayashta individual with a bachelor's degree, who is of average height and age. Similar conclusions would emerge from different assumptions.

Appendix Table C.4 presents similar regressions, using the ranking of the ad provided by the ad-placers as the dependent variable.[25] The results from these regressions are virtually identical to the ones presented in the previous table, as displayed graphically in Appendix Figures C.1 and C.2.

Appendix Tables C.7 and C.8 present similar regressions but this time exploring the determinants of which ad is selected by a letter writer or by another ad-placer or of the number of letters received by an ad-placer. In all these specifications, the importance of caste in the choice is at least as important as in the main specification. There are nevertheless interesting differences between these specifications and the ones presented here as far as the other variables are concerned, which we discuss in greater detail below.

4.3 Heterogeneity in preferences

The previous analysis suggests a strong horizontal preference for caste. The results presented in Appendix Table C.6 suggest that the preference for in-caste matching is not especially marked among the top 4 castes. Similar patterns were obtained when excluding the top 3 castes from the analysis. Thus, the estimates we presented above are not driven by a particular caste's desire for matching within their own group.

To further explore whether there is a lot of heterogeneity among ad-placers, a hierarchical binary logit model, as suggested by Rossi et al. (2006), was estimated using the parsimonious regression model above. This estimation method allows for the coefficients of our binary choice model equation to differ across individuals but imposes a normal distribution of heterogeneity. We allow the heterogeneity to depend on a few characteristics of the ad-placer, namely his or her caste, age, height and predicted income and the default prior suggested by Rossi et al. (2006). Figure 5 presents the results of this estimation for the preference for marrying within caste obtained using 20000 Markhov Chain Monte Carlo draws.[26]

The results suggest variations in this horizontal preference (over and above the preferences explicitly mentioned in the ads). Around one-third of the sample appears to have no preference for marrying within caste, a figure that is only slightly larger than the fraction of actual out-of-caste matches. As in the regressions presented above, the preference for in-caste matching is more marked among male than among female ad-placers, a group for which 40 percent do not have preference for marrying within caste. The mean preference for caste matching is only slightly smaller than the logit estimates found in Table 3, suggesting that being of the same caste increases the probability of responding to a letter by 15 percent, on average.

Furthermore, very similar results were obtained when we estimated the parsimonious regres-

[25] The sample size is a bit smaller due to missing observations (e.g., some ad-placers refused to provide ranking).
[26] The remaining estimates are available from the authors upon request.

sion using a OLS model but letting every single ad-placer have his or her own coefficient for the variable "same caste." This suggests that we have, as in the theoretical model above, a certain fraction of the population that appears to value endogamous matching much less than others.

4.4 Do these coefficients really reflect preferences?

We argue that these estimates provide us with information on the relative preferences for different attributes. There are two main objections to this interpretation which we examine here in detail.

4.4.1 Strategic behavior

A first concern is that ad-placers may behave strategically when they choose to which letters they will respond. For example, they may prefer not replying to a letter that appears to be "too good" because they think there is little chance of that relationship progressing. As we mentioned above, this is unlikely to be happening in this setting since the fact that the respondent has sent a letter to the ad-placer already signals his potential interest. Nevertheless, the issue is further investigated here.

We first compute an absolute measure of "quality" of the letter. To do so, we regress the probability that a letter in our sample is considered, without any interactions with characteristics of the ad-placer who received the letter. In other words, for P_{ij} a dummy indicating whether letter j is considered by ad-placer i, we estimate the equation $P_{ij} = X_j\beta + \epsilon_{ij}$ without any fixed effect for the ad-placer.

We form two versions of this indicator: with and without including the caste of the letter writer. The results presented here use those without caste but similar results were obtained with the caste variables included. The quality indicator is then given by $Q_j = X_j\hat{\beta}$. We also predict the quality of the ad-placer, using the same coefficients $Q_i = X_i\hat{\beta}$.

Figures 3 and 4 plot the probability of considering a letter based on the quality of the ad-placer and that of the letter. If the responses displayed strategic behavior, we would expect that low quality ad-placers would be less likely to consider high quality letters. In fact, Figures 3 and 4 show little difference in the relative probability of considering letters of different quality by the quantile of quality of the ad-placer, although higher quality ad-placers appear to consider on average a smaller fraction of letters of all quality levels. If anything, lower quality ad placers seem to respond to a higher fraction of higher quality respondents. Combining this with information about the letters received by each ad-placer's quality, this implies that the eventual number of letters considered are about evenly distributed across quality levels for ad-placers of the lowest quality and then become more and more skewed toward higher quality respondents for higher quality ad-placers.

Further evidence is provided by Appendix Table C.5 where similar regressions as the ones presented above are presented but this time restricting the sample to letters where the quality of the ad-placers and the quality of the letter writers are relatively close. The behavior of the ad-placer seems to be fairly similar when looking at the overall sample compared to this relatively lower quality sample, either in terms of considering letters or ranking them.

Interestingly, the decision to respond to an ad (reported in the appendix tables) seems to reflect more strategic behavior than the choice of whether to respond to a letter an ad-placer received. For example, in the decision of whether an ad-placer replies to another ad, and in the decision of whether a letter writer replies to another ad (Appendix Table C.7 and C.8), education loses its previous importance and appears to potentially decrease one's attractiveness. Finally, when we regress the number of responses received on a polynomial function of our measure quality Q_i (computed as before), we find that the best fit between quality of an ad and the overall number of responses is an inverse-U shaped curve. This may indicate that, at the ad stage, higher quality ads are only replied to by people who stand a chance.

Thus, there is evidence that families behave strategically when they respond to ads but much less so subsequently. This is perhaps not surprising, as they have to choose between a very large number of ads. While the average person sees more than 800 ads every Sunday over the 12 months they spend on the market before getting married, they only respond to on average 16 of these for females and 35 for males. In contrast, it appears that each ad-placer considers each of the 40 letters received during their search as a potential prospect, and therefore do not behave strategically about whom to respond to (ad-placers respond to about 30 percent of the letters they receive).[27]

4.4.2 What does caste signal?

One of our main empirical results is the fact that families (ad-placers as well as people who write to them) are much more likely to write to, and to follow up with, people from their own caste. Caste preferences thus display a strong horizontal component. Does this reflect a preference for caste in itself, or does caste signal something else?

We first explore the possibility that caste is a shortcut for many variables, perhaps unobserved by the ad-placer and us, but reflecting a prospective spouse's background and culture. People would then match within their caste to marry people like them. However, the strong preference for caste does not seem to be affected by controlling for a host of variables including cultural variables (e.g., ability to sing, which is often mentioned in the ads as a desirable characteristic of women) and it remains very strong in regressions restricted to the four highest castes, who are

[27]This is less costly than an equilibrium where letter writers would send a message to most ads and would leave the ad-placers to strategically consider (or not) the letters received.

culturally and economically more homogenous than the rest (Appendix Table C.6). It therefore does not appear that caste is just a proxy for cultural similarity. Furthermore, columns (3) and (8) of Table 3 also include a dummy variable for being from the same broad caste group. The results suggest that it is the narrow caste that matters for preference. If caste was a proxy for cultural identity, broad caste groupings should be stronger than smaller groups.

A second possibility is the preference of ad-placers for letter writers who are from the same caste as themselves reflects the fact that, in equilibrium, only people with some bad unobservable characteristics write to people who are not in their castes (or who are above them or below them). Writing "out of caste" would then be a signal of bad quality.

We first look at whether people who write to, or receive letters from, people belonging to other castes are observationally different from those who do not. In columns 1 and 3 of Panel A in Table 4, we show the average quality index Q for ad placers who told us that they have responded to at least one letter from a caste that is below or above them, compared to the quality of those who only responded to people from their caste. Each cell is the difference in mean quality between those who satisfy the condition and those who do not. This table indicates that there does not seem to be significant observable differences between people who write to someone from a different caste and people who do not. There is also no difference between the people who receive letters from other castes, and those who don't (panel B).

This still leaves open the possibility that these individuals are different along unobservable dimensions. However, we have an excellent measure of the unobservable (at the time of ad placing or letter writing) quality of people: we know their eventual outcome. We compute our quality index for each ad-placer's future spouse, and we contrast the eventual marriage outcomes of those who have written to at least one person from another caste to that of people who have only written to people within their caste. In an alternative specification, we also regress the quality of the eventual mate of an ad-placer on the share of ads they replied to that were not from the same caste. The results (presented in Columns 2 and 4 of Table 4) suggest that the ultimate marriage outcome of those who write out of caste are no different than those who do not (panel A). Likewise, those who get letters from other castes eventually marry people of the same observable quality (panel B). This is a strong indication that writing out of caste does not send the signal that something is "wrong" with the ad-placer.

These results therefore suggest that the fact that ad-placers are more likely to follow up with people from their own caste reflects a true preference for eventually marrying within the same caste. This preference seems to be related to caste itself, rather than characteristics caste could be a proxy for. Compared to the other attributes, this preference also appears to be extremely strong: it appears that the parents of prospective grooms or brides would be willing to give up a lot to ensure that their child marries within their caste. Furthermore, the preference for caste

appears to be strongly "horizontal" rather than "vertical,"as defined above in the theoretical section.

4.5 Do these preferences reflect dowry?

We have so far ignored dowries. We argued in the theory section that even if some people do eventually ask for dowries, the decision of who to write back to will be based on people's true (i.e, not dowry-based) preferences, as long as the cost of pursuing the option until the information on dowry –or other unobservable variable– is revealed, is not too high. One way to check the validity of this argument is to test one of its implications: those who either say that they do not want dowry should be treated the same as others. To verify this conjecture in the data we re-estimate the preferences in the sample of letters that explicitly mention not wanting a dowry. In Table 5 we interact not wanting a dowry with each characteristic of the letter. The full specification is presented in columns (1) and (2), and the parsimonious specification is presented in columns (3) and (4).[28] The even columns correspond to the interaction terms and the odd columns to the main effect. The results are noisier for the interactions than for the main effects given the sample size, but, overall, we cannot reject that the interaction terms are jointly equal to zero.

Interestingly, caste plays an even bigger role for this sample (the coefficient of the interaction between not wanting a dowry and being of the same caste is positive, although it is not significant), while the role of predicted income does not change. This suggests an even larger marginal rate of substitution between caste and income, which is the opposite of what would have been predicted if rich grooms were also thought to require higher dowries. In that case, for grooms who state they will not demand a dowry, income would become more valuable while we find the opposite pattern to hold.

In addition, we find that ad-placers who either announce that they will not offer a dowry or state that they will not demand one do not receive systematically different numbers of letters, and their attributes as mentioned in the letter are valued similarly. Also the quality of their responses and their eventual match is not significantly different than others, except for female ad-placers who receive slightly worse applicants when they announce that they would not offer a dowry. The results are not reported to save space, but are available from the authors.

[28] We present these results only for the "bride-wanted" sample since only prospective grooms specify whether or not they will accept a dowry. No prospective bride is advertised as refusing to pay a dowry in the letters and a very small proportion do so in the ads.

5 Stable matching estimates

Having established that strong horizontal caste preferences in our sample exist, we compute the set of stable matches implied by the preferences estimated to further study the role of caste in equilibrium. A stable match is defined, following Gale and Shapley (1962), as a pairing where nobody who is matched would rather be with another partner who would also prefer being with them (see Hitsch et al. 2006 and Lee 2007 for other applications of this method to the marriage market). These simulated matches will then be used to answer questions regarding the equilibrium role of caste.

5.1 Empirical strategy

The pool of men and women attempting to match within this market is defined as the entire set of ads posted during the period of the survey, from October 2002 to March 2003 (most individuals on the market usually place one and only one ad, which makes this approximation acceptable).

We want to construct ordinal preferences over the entire set of bride (groom) wanted ads for each man (woman), in the sample. To do so we use the estimated parameters in equation (8) to construct the predicted "utility" that each man i in the sample (the set of ads) would get from matching with woman j (and vice versa for women) using the following equations.[29]

$$U_{ij}^{k} = \hat{\alpha}_k X_i + \hat{\beta}_k f\left(X_i, X_j\right) \text{ for } k = m, f \tag{10}$$

Functions U^m and U^f are then transformed into ordinal ranking such that

$$R_{ij}^{k} = n \quad if \quad \left\{ \begin{array}{c} U_{ij'}^{k} > U_{ij}^{k} > U_{i\tilde{j}}^{k} \\ and \quad R_{ij'}^{k} = n-1 \quad and \quad R_{i\tilde{j}}^{k} = n+1 \end{array} \right\} \text{ for } k = m, f.$$

The preference estimates for the results presented below were all obtained from the linear specification as presented in columns (1) and (6) of Table 3. However, extremely similar results were obtained using the logit specification or the ranking estimates as presented in Appendix Table C.4. Applying this methodology for all males and females in the sample, this generates a full set of ordinal preferences for each ad-placer with respect to all ad placers of the opposite gender.

The Gale-Shapley algorithm can be computed in many ways. In most of the results presented in this section, we assume that men make an offer to women. We later explore how the results

[29] The input required by the stable matching algorithm is a measure of ordinal and not cardinal utility, so fixed effects can be ignored. This is because the fixed-effect of male i, for example, simply affects the overall preference of person i towards all potential mates and not the relative ranking of each mate within his set of preferences.

change when women propose to men instead. When men propose to women, the algorithm works as follows. All men first propose to their most highly-ranked women. Women consider all the offers they receive and select the best one (staying single is considered to be a worse option than any marriage). All men who haven't been retained then select their second choice. If a woman receives a new offer that is preferable to the one she is currently holding, she releases the old offer and this man must then propose to the next woman on his list. This continues until all men have been matched. Since they are the long side of the market, some women will remain single. Ties are broken randomly, without loss of generality in this setting (unlike the example discussed by Erdil and Ergin 2008).

In order to obtain confidence intervals for the results of the matching algorithm, 1000 estimates of the parameter estimates of equation (9), α and β were obtained by bootstrapping the above estimation procedure.[30] Then, using each of the 1000 sets of parameters, the matching algorithm was separately run. This resulted in 1000 stable matches that define the range of outcomes that could stem from the distribution of preference parameters. All the stable matching results will present the 2.5th and 97.5th percentiles of each characteristic of interest to bound the range of results obtained.

One may worry that the assumption of frictionless matching, implied by the Gale-Shapley algorithm, is inappropriate. To explore this issue, we introduce search frictions in the following way. First, we constrain males to contact individuals close to their unconstrained optimal choice (within 1000 ranks). Second, at every offer period, a man may be unable to offer to a particular woman with 75 percent probability and may thus be constrained to skip this woman and offer to the next preferred candidate. With search frictions, some males remain unmatched but without all find a spouse because they are on the short-side of the market.

Finally, to compare the results of the algorithm to those observed in the data, the summary statistics for the algorithm results are computed only for the individuals in our original interview-sample, though using the ad-placer samples gives very similar results.

5.2 Results

This section presents the stable matches estimated with the algorithm as described above. We ask two distinct questions: who finds a spouse, and who marries whom. We then compare the simulated outcomes to the actual ones.

[30] This was done using a "block bootstrap" by ad-placer, that is, either all letters in response to an ad are randomly selected into the sample or they are all excluded.

5.2.1 Who stays single?

In Table 6 we show the mean differences in the value of key attributes between single and married females in the simulations and in the observed data, that is, the difference between the characteristics of single women and those who are married. Columns 1 and 2 show the 2.5 percent and 97.5 percent of the distribution of these differences within the algorithm. Column 3 presents the mean differences in the actual sample with the 95 percent confidence interval around that mean shown in columns 4 and 5.

In most cases, the differences between those who get married and those who stay single observed in the stable matching have the same signs as the actual differences. Older, shorter, darker skinned, less beautiful and less educated women are more likely to be single in both the stable matches and the actual data. Commerce graduates are also less likely to be single. Some attributes had no effect on the probability of being married or single, including being from West Bengal, being beautiful or very beautiful, and occupational wage and income reported in the ad. For seven out of the sixteen variables, the actual difference between single and married in our data lies within the confidence interval of the stables matches. In five more cases, the confidence intervals overlap.[31]

There are two variables for which the stable matching algorithm gets the sign wrong. The most important one is the role of caste.[32] While we predict that the singles would be of a lower caste than those who are married, it is not true in the real data, where the singles are, if anything, of slightly higher castes.

In most cases where the point estimate of the difference in the actual data does not lie within the bounds of the stable matches estimate, the stable matches overestimate the differences between the variable. This probably reflects the fact that factors other than these attributes eventually determine whether or not people decide to marry: this will thus dampen the role of the variable in the case of actual matches.

As a first pass to investigate this possibility, panel B introduces search frictions. They do not appear to affect the results very much. The resulting characteristics of married and single females are quite similar in both scenarios, and the fit with the real matches is not improved much.

Panel C repeats the exercise for males. Since men are on the short side of the market, without any search frictions, all men will be married. The algorithm results are thus only presented in the case of search frictions. The signs are now congruent for all the variables, and the observed mean differences between those who stay single and those who get married fits within the 95 percent

[31] However, because some characteristics are clearly outside the overlapping region, a chi-square test of equivalence of the moments of the algorithm with the mean values observed in the actual match data rejected their equivalence.

[32] The other one being whether a woman has a science degree.

predicted by the stable matching algorithm in eight out of thirteen characteristics (although it is to be acknowledged the algorithm does not produce very tight predictions). Moreover the main characteristics have the expected effect on the probability of being married: men who are more educated, have a science degree, and report higher income or wages, are less likely to remain single, both in reality and and in the results of the matching algorithm.

5.2.2 Who marries whom?

We now compare the characteristics of the couples in the stable matches and in our actual sample. Table 7 displays the main results. Columns 1 and 2 present the lower and upper bound for the stable matches, using the "considered" response to estimate the preferences while columns 3 to 5 present the actual comparison between ad-placers and the letters they consider. Columns 6 to 8 compares the ad-placers and their actual matches. All the differences are expressed in terms of the difference between the husband and the wife.

The stable matching algorithm predicts the characteristics of the couples reasonably well. For all the statistics we look at, the sample equivalent in the actual marriages fits within the range of the stable matches estimate in 14 cases out of 21, and the confidence intervals overlap in 15 cases, even though for many variables, the bounds on the stable matches are quite tight.[33]

Not surprisingly, a dominant feature is the tendency to marry within one's caste. The stable matching predicts that 87 to 97 percent of the couples will have the same caste. In practice, a lower share (almost 70 percent) of the couples are from the same caste.

Turning to other characteristics, the prediction regarding age are roughly similar in the simulations and in the data. Husbands are almost six years older than their wives on average. Height differences are slightly underestimated but we predict too much assortative matching by height as given by the spousal heights correlation. Both the data and the simulations suggest that husbands are 10 to 12 centimeters taller than their wives.

For education, we correctly predict the fraction of couples with the same education level and the correlation between the education of the spouses, although we tend to predict that husbands will be less educated than their wives, and the opposite is true in the data. This is surprising, and probably comes from the fact that men from the top of the educational distribution may be less likely to report their education than females as they can signal that quality using their wage/occupation.

Comparing our indices of quality, we find that males have higher indices than their spouses though this measure is slightly overestimated compared to the observed data.[34] These indices

[33]However, because the stable matching differs greatly from the actual matches in a few instances, a chi-square test of the algorithm moments and the mean values for either considered or match individual rejected the hypothesis of their equality.

[34]This is driven by two elements. First, male letter writers have higher response rates and thus the indices are

are also positively correlated according to the algorithm and in reality.

The algorithm does not have much to say on predicted wage and income differences. This appears to stem from the fact that few women report their wage and income and that these variables are not part of the estimated preferences for males. Finally, we seem to severely overestimate the correlation in family origins.

Introducing search frictions slightly improves the fit of the algorithm result. Although the results are not altered greatly, they are modified in a way that generally increases their resemblance to the observed data. The education and wage differences become more positive with search frictions than without them. Height differences are now including the observed data in the case where considered probabilities are used as preference parameters. Family origin matching is still overestimated when compared to the observed matches. Still, the imposition of these fairly substantial search frictions has limited impact on the results.

We also computed the equilibrium under two variants, presented in Table C.9. First, we computed the equilibrium under the assumption that women propose rather than men. The equilibrium we obtain is very similar in terms of who marries whom. Actually, less than 2 percent of the matches differ between the two algorithms.[35] Furthermore, while not shown, the characteristics of who remains single and who finds a match are almost identical when women propose and a very small number of women (less than 0.025 percent) are single when they propose and find a spouse when men propose. This suggests an almost unique stable matching. Finally, we also imposed a balanced sex ratio by randomly selecting a subset of females equal to the number of male ads in the sample. While this creates some differences in the algorithm, the results are still fairly similar to the ones presented in the main tables.

6 The role of caste preferences in equilibrium

6.1 Model Predictions

In Section 4, we saw that there was a strong preference for marrying within one's caste. Men were willing to sacrifice up to four categories of education and women more than 300 percent of a man's income in order to marry within one's caste. We also saw that, indeed, about 70 percent of the marriages take place within caste. While individuals appear to be ready to pay a high price to marry within their caste, do they end up paying for it in equilibrium? More generally, does the preference for marrying within caste affect other dimensions of matching?

larger for males than for females in general. Second, since women with lower quality indices are remaining single, the matches are such that there is an even larger difference between spouses.

[35] This is similar to findings by Roth and Peranson (1999) in the context of medical residency matching and by Pathak and Sönmez (2008) in the context of Boston public school matching.

In Section 2, the theoretical model emphasized that the equilibrium role of caste crucially depends on whether preferences for caste are horizontal or vertical. Section 4 has then argued that the estimation of preferences suggests that the preference for caste is horizontal rather than vertical.

The theoretical model discussed above also suggests that one important element is whether the distribution of male and female "quality" is balanced across castes. In our sample, we know that there is a surplus of females given that more ad-placers are looking for a groom. However, is there evidence of a difference in the quality distribution across castes that differ by gender? To evaluate this question, we used the "quality" measure defined above (without any caste parameter) and compared the overall distribution of quality by caste for males and females among the interview sample. We find that the distributions are fairly similar for all major caste groups (Brahmin, Kayastha, Baisya and Sagdope), but are less similar for caste groups with fewer observations. These results hold whether one compares the distribution in quality among the interview sample or the letter sample.

Finally, the model we elaborated earlier also suggests that the equilibrium price will be low when there is a group that does not have caste preferences. We find that in our data between 25 to 30 percent of individuals are willing to marry outside their caste. This roughly corresponds to the number of matches observed that are not within one's caste, although not all individuals who say they would be willing to marry outside their caste eventually do so (and vice versa).

Given this evidence, we should expect a relatively low impact of caste on the pattern of matches along other dimensions.

6.2 Simulations

What do the algorithm results tell us about the actual role of caste in the matching equilibrium? Table 8 takes one cut at this issue. The first columns of panel A of Table 8 reproduce columns 1 and 2 of the first panel of Table 7. The second panel constrains all marriages to take place within one's caste. Panel C entirely ignores caste when computing the preference of each ad-placer for each prospective bride or groom.

The striking result in this table is that neither of these manipulations greatly affects how matches look like along non-caste dimensions. As expected, the correlations in age, height and education increase as the preferences for caste diminishes (they are the highest when matches are restricted to be within caste, and the lowest when preferences for caste is "shut down"), but the gradient is fairly low, and very few of the other variables are affected.

Moreover, the proportion of within-caste marriage falls by a large fraction when preferences are caste-blind. This suggests that caste does not proxy for other attributes. There are many potential matches for each person, both within and outside his or her caste.

Columns (3) to (10) present the algorithm results by key caste groups. These results suggest that the conclusions drawn above are fairly similar across caste groups, despite the fact that the sub-castes within the Baisyas and the Sadgopes are relatively smaller than those within the Brahmins. However, imposing caste-blindness appears to more importantly affect smaller castes than Brahmins or Kayashtas. Some correlations among the Sagdopes, in particular age and education correlations, appear to fall once one imposes within-caste matching.

Overall it seems that once the algorithm imposes caste-blindness, the individuals marry almost identical individuals but from another caste. This would suggest that the equilibrium price of caste ought to be low. To further study this pattern, we look at the actual matching patterns of our sample. We found no evidence that men or women who marry outside their caste sacrifice "quality" measured in a variety of ways. However, this could be due to selection. That is, individuals who have less of a preference for caste would select to marry outside their caste. Since their "cost" of caste matching is lower, this is what we would measure in equilibrium.

Therefore, we turn to the results of the algorithm to attempt to alleviate this concern since in this context there are no unobservable determinants of taste. The conceptual exercise here consists in comparing the spouses of two observationally equivalent individuals where one is matched within his or her caste and the other is not. To do this, a regression controlling for all of the ad-placer's characteristics correlated various measures of quality of the match with an indicator of whether the match is within or outside one's caste. Such regressions were run for each iteration of the algorithm, and Table 9 presents the mean and the 2.5 and 97.5 percentile of the distribution of the coefficients on whether or not the couple was within the same caste. These results suggest that prices of matching within caste are small, insignificant, and often in the wrong direction. For example, individuals who marry within their own caste are also more likely to marry more educated individuals.

On the other hand, our theory would make us expect that vertical characteristics, such as education, should have a positive price; if we did not find this, the evidence on the low price of caste would be less compelling. We therefore compute the equilibrium price of education in a similar fashion. The left hand panel of Table 9 suggests that as opposed to caste, individuals are forced to choose between, for example, beauty and the educational level of a woman. A man who marries a woman who has more education also marries one who is older, less beautiful and darker-skinned. Little correlation is found between a prospective groom's education and other qualities.

We thus find that the equilibrium price of caste is very small and that altering the way caste is perceived by individuals does not transform the overall matching equilibrium significantly. This is consistent with our theoretical model and the estimated preferences we obtained in the context where preferences for caste are horizontal.

7 Conclusion

Our results indicate that while caste is highly valued in terms of preferences, it does not require a very high price in equilibrium. This is consistent with assuming that preferences are relatively horizontal and that the populations are close to being balanced. Both these conditions appear to hold in the data we collected for arranged marriages in West Bengal.

A number of conclusions follow from this. First, there is no reason to expect that economic growth by itself will undermine caste-based preferences in marriage. Second, caste-based preferences in marriage are unlikely to be a major constraint on growth. Finally, one might worry that if caste becomes less important inequality might increase along other dimensions as we will see more assortative matching. Given that the matching is already close to being assortative this is probably not an important concern.

Despite the value placed on caste and its low equilibrium price, 30 percent of people in our sample *do not* marry within their caste. They apparently do not gain much by marrying out of caste. So why do they do it? In part, this comes from heterogeneity in caste preferences, with some people having caste-neutral preferences. But there is something else. A substantial fraction of the marriages that are not within caste are "love marriages." About 40 percent of the sons and daughters of our respondents eventually marry through a channel other than the ads, and 20 percent enter into a "love marriage." So the institution that economic forces are not able to destroy may be endangered by love.

References

Anderson, S. (2003). Why Dowry Payments Declined with Modernization in Europe but Are Rising in India. *The Journal of Political Economy 111*(2), 269–310.

Bargain, O., S. K. Bhaumik, M. Chakrabarty, and Z. Zhao (2007). Returns to education and earnings differences between Chinese and Indian wage earners. Available at `http://www.iza.org/conference_files/worldb2007/bargain_o1569.pdf`.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy 81*(4), 813–46.

Bose, N. K. (1958). Some aspects of caste in Bengal. *The Journal of American Folklore 71*, 397–412.

Cole, H. L., G. J. Mailath, and A. Postlewaite (1992). Social norms, savings behavior, and growth. *The Journal of Political Economy 100*(6), 1092–1125.

Dugar, S., H. Bhattacharya, and D. Reiley (2008). Can't buy me love: A field experiment exploring the tradeoff between income and caste in the Indian matrimonial market. Unpublished manuscript available at http://www.aeaweb.org/annual_mtg_papers/2009/retrieve.php?pdfid=170.

Erdil, A. and H. Ergin (2008). What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review 98*, 669–689.

Fernandez, R. (2003). Household formation, inequality, and the economy. *Journal of the European Economic Association 1*, 683–697.

Fernandez, R. and R. Rogerson (2001). Sorting and long run inequality. *Quarterly Journal of Economics 116*, 1305–1341.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics 121* (2), 673–697.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2008). Racial preferences in dating. *Review of Economic Studies 75*(1), 117–132.

Foster, A. and M. Rosenzweig (2001). Missing women, the marriage market and economic growth. Unpublished manuscript available at http://adfdell.pstc.brown.edu/papers/sex.pdf.

Fox, J. (2007). Estimating matching games with transfers. Unpublished manuscript available at http://home.uchicago.edu/~fox/_Media/foxmatching-2.pdf.

Gale, D. and L. S. Shapley (1962). College admissions and the stability of marriage. *The American Mathematical Monthly 69*(1), 9–15.

Glinskaya, E. and M. Lokshin (2005). Wage differentials between the public and the private sectors in India. World Bank Policy Research Working Paper No. 3574 available at http://ssrn.com/paper=719121.

Hitsch, G. J., A. Hortacsu, and D. Ariely (2006). What Makes You Click? Mate Preferences and Matching Outcomes in Online Dating. MIT Sloan Research Paper No. 4603-06 available at http://ssrn.com/paper=895442.

Lam, D. (1988). Marriage markets and assortative mating with household public goods: Theoretical results and empirical implications. *The Journal of Human Resources 23*(4), 462–487.

Lee, S. (2007). Preferences and choice constraints in marital sorting: Evidence from Korea. Unpublished manuscript available at http://soohlee.googlepages.com/soohyunglee_JMP.pdf.

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review 96*(4), 1225–1252.

Murphy, K. M. and R. H. Topel (1985, October). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics 3*(4), 370–79.

Park, M. (2007). M&a incentives and outcomes: Evidence from the mutual fund industry. Unpublished manuscript available at http://www.econ.umn.edu/~mpark/Merger%20Paper.pdf.

Pathak, P. A. and T. Sönmez (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. Forthcoming, *American Economic Review*.

Risley, H. H. (1981). *The Tribes and Castes of Bengal*, Volume II. Firma Mukhopadhyay.

Rosenzweig, M. R. and O. Stark (1989). Consumption smoothing, migration, and marriage: Evidence from rural India. *The Journal of Political Economy 97*(4), 905–926.

Rossi, P., G. Allenby, and R. McCulloch (2006). *Bayesian Statistics and Marketing*. Wiley.

Roth, A. E. and E. Peranson (1999, September). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review 89*(4), 748–780.

8 Tables and figures

Table 1: **Summary statistics:Ad-placers**

Variable	Ads placed by females				Ads placed by males			
	Full set (N=14172)		Interviewed (N=506)		Full set (N=8038)		Interviewed (N=277)	
	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.
Number of responses			22.67	19.84			82.71	76.10
Caste								
Brahmin	0.26	0.44	0.26	0.44	0.27	0.44	0.25	0.44
Baidya	0.04	0.20	0.04	0.20	0.03	0.18	0.05	0.21
Kshatriya	0.02	0.13	0.02	0.13	0.02	0.13	0.01	0.12
Kayastha	0.30	0.46	0.35	0.48	0.29	0.45	0.32	0.47
Baisya and others	0.18	0.39	0.19	0.39	0.20	0.40	0.18	0.38
Sagdope and others	0.13	0.34	0.10	0.30	0.13	0.34	0.12	0.33
Other castes	0.02	0.14	0.02	0.13	0.02	0.12	0.03	0.16
Scheduled castes	0.06	0.23	0.03	0.16	0.05	0.21	0.04	0.20
Physical characteristics								
Age	26.68	3.90	26.59	3.65	31.58	4.31	32.14	4.45
Height (meters)	1.56	0.04	1.58	0.04	1.68	0.06	1.70	0.06
Skin tone	2.36	0.84	2.30	0.80				
Very beautiful	0.06	0.24	0.08	0.27				
Beautiful	0.56	0.50	0.44	0.50				
Education and Income								
Less than high school	0.03	0.16	0.02	0.15	0.01	0.12	0.01	0.08
High school	0.06	0.23	0.08	0.28	0.07	0.25	0.08	0.27
Post-secondary	0.01	0.10	0.00	0.04	0.03	0.18	0.04	0.20
College	0.46	0.50	0.49	0.50	0.36	0.48	0.35	0.48
Master's	0.29	0.45	0.26	0.44	0.17	0.37	0.15	0.36
PhD	0.06	0.24	0.05	0.22	0.13	0.34	0.18	0.39
Other degree	0.00	0.04	0.01	0.10	0.01	0.08	0.01	0.10
Humanities/Arts	0.66	0.47	0.58	0.49	0.12	0.33	0.05	0.21
Commerce	0.11	0.31	0.12	0.33	0.37	0.48	0.40	0.49
Science	0.28	0.45	0.30	0.46	0.55	0.50	0.55	0.50
Other field	0.01	0.11	0.01	0.07	0.02	0.15	0.00	0.00
Log wage	5.55	0.36	5.54	0.35	5.20	0.79	5.61	0.53
Log income	9.22	0.83	8.75	0.77	9.46	0.75	9.44	0.67
Geography								
Living in Calcutta	0.51	0.50	0.80	0.40	0.50	0.50	0.76	0.43
Family from West Bengal	0.44	0.50	0.39	0.49	0.45	0.50	0.39	0.49
Demands mentioned								
Only within caste	0.09	0.29	0.10	0.30	0.10	0.30	0.08	0.28
Caste no bar	0.31	0.46	0.33	0.47	0.26	0.44	0.24	0.43
No dowry demanded	0.03	0.16	0.02	0.12	0.12	0.32	0.10	0.31
Ads which omit...								
Caste	0.02	0.13	0.00	0.04	0.03	0.16	0.01	0.08
Age	0.01	0.10	0.01	0.12	0.02	0.13	0.04	0.20
Height	0.04	0.19	0.04	0.19	0.10	0.30	0.11	0.31
Education	0.10	0.30	0.08	0.27	0.22	0.42	0.18	0.39
Field	0.27	0.44	0.25	0.43	0.39	0.49	0.30	0.46
Residence	0.86	0.35	0.84	0.37	0.70	0.46	0.52	0.50
Family origin	0.29	0.45	0.23	0.42	0.32	0.47	0.29	0.45
Wage	0.83	0.38	0.84	0.37	0.25	0.43	0.57	0.50
Income	0.98	0.13	0.97	0.16	0.78	0.41	0.74	0.44
Skin tone	0.23	0.42	0.21	0.41				
Beauty	0.25	0.43	0.27	0.44				

Statistics are computed only among individuals reporting a given characteristic

Table 2: **Summary statistics:Letters and matches**

Variables	Ads placed by females				Ads placed by males			
	Letters (N=5630)		Matches (N=158)		Letters (N=3944)		Matches (N=131)	
	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.	Mean	Sd. Dev.
Considered	0.34	0.47			0.28	0.45		
Caste								
Brahmin	0.23	0.42	0.27	0.45	0.21	0.41	0.24	0.42
Baidya	0.03	0.17	0.04	0.19	0.04	0.19	0.05	0.23
Kshatriya	0.01	0.10	0.01	0.08	0.02	0.14	0.03	0.17
Kayastha	0.38	0.48	0.43	0.50	0.36	0.48	0.37	0.49
Baisya and others	0.20	0.40	0.15	0.36	0.20	0.40	0.16	0.37
Sagdope and others	0.12	0.32	0.07	0.26	0.11	0.32	0.11	0.31
Other castes	0.01	0.08	0.01	0.11	0.02	0.14	0.01	0.09
Scheduled castes	0.04	0.19	0.02	0.14	0.04	0.19	0.03	0.17
Same caste	0.66	0.47	0.68	0.47	0.64	0.48	0.72	0.45
Difference in caste	-0.17	1.37	0.10	1.43	-0.04	1.23	-0.11	1.08
Physical Characteristics								
Age	32.60	4.37	32.49	3.67	26.34	3.96	27.33	3.67
Age difference	6.25	2.92	6.61	2.95	5.93	2.65	4.60	2.84
Height (meters)	1.70	0.06	1.71	0.08	1.58	0.04	1.59	0.05
Height difference (m)	0.12	0.06	0.13	0.08	0.12	0.07	0.12	0.06
Skin tone					1.41	0.77		
Very beautiful					0.10	0.31		
Beautiful					0.51	0.50		
Education and Income								
Less than high school	0.00	0.06	0.00	0.00	0.02	0.12	0.01	0.09
High school	0.08	0.27	0.06	0.22	0.16	0.37	0.08	0.28
Post-secondary	0.04	0.19	0.03	0.16	0.00	0.06	0.02	0.12
College	0.51	0.50	0.35	0.48	0.58	0.49	0.44	0.50
Master's	0.21	0.41	0.25	0.44	0.18	0.39	0.34	0.48
PhD	0.13	0.33	0.32	0.47	0.02	0.13	0.11	0.32
Other degree	0.03	0.18	0.00	0.00	0.04	0.19	0.00	0.00
Same education level	0.44	0.50	0.42	0.49	0.37	0.48	0.46	0.50
Male is more educated	0.28	0.45	0.45	0.50	0.44	0.50	0.23	0.42
Humanities/Arts	0.13	0.33	0.52	0.50	0.63	0.48	0.79	0.41
Commerce	0.34	0.47			0.11	0.31		
Science	0.51	0.50	0.48	0.50	0.25	0.43	0.21	0.41
Other field	0.02	0.14	0.00	0.00	0.01	0.12	0.00	0.00
Log wage	5.47	0.59	5.53	0.57	5.50	0.35	5.46	0.36
Log income	9.31	0.73	9.47	0.79	8.85	0.68	1.75	3.54
Geography								
Living in Calcutta	0.55	0.50	0.59	0.50	0.54	0.50	0.53	0.50
Same residence	0.50	0.50	0.64	0.49	0.44	0.50	0.42	0.50
Family from West Bengal	0.39	0.49	0.46	0.50	0.41	0.49	0.42	0.50
Same family origin	0.75	0.43	0.75	0.43	0.71	0.46	0.72	0.45
Demands mentioned								
No dowry demanded	0.07	0.26	0.00	0.00				
Letters which omit								
Caste	0.30	0.46	0.01	0.11	0.28	0.45	0.02	0.12
Age	0.04	0.20	0.00	0.00	0.03	0.17	0.00	0.00
Height	0.13	0.33	0.00	0.00	0.08	0.27	0.00	0.00
Education	0.08	0.27	0.00	0.00	0.04	0.19	0.00	0.00
Field	0.20	0.40	0.39	0.49	0.25	0.43	0.22	0.42
Residence	0.15	0.36	0.00	0.00	0.19	0.40	0.00	0.00
Family origin	0.31	0.46	0.03	0.18	0.27	0.44	0.00	0.00
Wage	0.44	0.50	0.08	0.28	0.86	0.35	0.79	0.41
Income	0.66	0.47	0.31	0.46	0.98	0.14	0.04	0.19
Skin tone					0.14	0.35	1.00	0.00
Beauty					0.36	0.48	1.00	0.00

Statistics are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Statistics are computed only among individuals reporting a given characteristic. Ads placed by females (males) received letters by males (females): the first four columns refer to prospective and actual grooms, the last four to prospective and actual brides.

Table 3: **Probability of considering a letter**

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same caste	0.1317***		0.1347**	0.1395***	0.8604***	0.1707***		0.1769***	0.1800***	1.0454***
	(0.0329)		(0.0425)	(0.0330)	(0.2068)	(0.0351)		(0.0442)	(0.0352)	(0.2052)
Same main caste			0.0273					-0.0331		
			(0.0485)					(0.0554)		
Diff. in caste*Higher caste male	-0.0119		-0.0276	-0.0108	-0.0788	-0.0175		-0.0099	-0.0138	-0.1990
	(0.0151)		(0.0197)	(0.0152)	(0.0928)	(0.0170)		(0.0232)	(0.0171)	(0.1081)
Diff. in caste*Lower caste male	0.0145		0.0056	0.0103	0.1393	-0.0399*		-0.0301	-0.0428*	-0.2958**
	(0.0133)		(0.0160)	(0.0134)	(0.0903)	(0.0172)		(0.0220)	(0.0173)	(0.0990)
Same caste*only within	0.0954		0.0918	0.0968	35.1982	0.1234		0.1217	0.1162	1.5756
	(0.1093)		(0.1093)	(0.1097)	(1288.88)	(0.1409)		(0.1410)	(0.1418)	(1.7103)
Diff. in caste*only within	-0.0163		-0.0158	-0.0188	-11.6502	0.0024		0.0010	-0.0056	0.0674
	(0.0400)		(0.0400)	(0.0402)	(429.6274)	(0.0596)		(0.0596)	(0.0597)	(0.6857)
Same caste*no bar	-0.0560		-0.0549	-0.0563	-0.4950*	-0.0565		-0.0574	-0.0629	-0.2599
	(0.0366)		(0.0366)	(0.0367)	(0.2187)	(0.0428)		(0.0429)	(0.0430)	(0.2424)
Diff. in caste*no bar	-0.0084		-0.0098	-0.0052	-0.0528	0.0121		0.0118	0.0115	0.1194
	(0.0121)		(0.0121)	(0.0121)	(0.0786)	(0.0151)		(0.0152)	(0.0152)	(0.0880)
Diff. in age	-0.0019	-0.0035	-0.0019	-0.0032	0.1647***	0.0443***	0.0471***	0.0443***	0.0394***	0.2933***
	(0.0047)	(0.0047)	(0.0047)	(0.0047)	(0.0458)	(0.0083)	(0.0083)	(0.0083)	(0.0082)	(0.0545)
Squared diff. in age	-0.0008**	-0.0008**	-0.0008**	-0.0008**	-0.0203***	-0.0023***	-0.0025***	-0.0023***	-0.0023***	-0.0150***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0035)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0038)
Diff. in height	1.2508***	1.3455***	1.2490***	1.3028***	8.1805***	0.7228*	0.6829*	0.7153*	0.7585*	10.2634***
	(0.2745)	(0.2754)	(0.2745)	(0.2752)	(1.7128)	(0.3329)	(0.3348)	(0.3331)	(0.3339)	(2.6758)
Squared diff. in height	-3.4695***	-3.8398***	-3.4465***	-3.5684***	-22.4174***	-6.2532***	-6.1518***	-6.2375***	-6.3265***	-60.1849***
	(0.9692)	(0.9718)	(0.9694)	(0.9709)	(5.9882)	(1.2451)	(1.2522)	(1.2455)	(1.2491)	(10.2198)
High school	0.0732	0.0907	0.0751		0.0770	0.1043	0.1133	0.1038		0.6122
	(0.1097)	(0.1102)	(0.1097)		(0.6478)	(0.0623)	(0.0628)	(0.0624)		(0.3896)
Post-secondary	0.1216	0.1413	0.1238		0.3391	0.0832	0.0701	0.0808		0.5283
	(0.1187)	(0.1192)	(0.1188)		(0.6995)	(0.1403)	(0.1409)	(0.1403)		(0.8193)
Bachelor's	0.1019	0.1132	0.1024		0.2708	0.0966	0.1224	0.0965		0.3744
	(0.1183)	(0.1188)	(0.1183)		(0.6942)	(0.0879)	(0.0884)	(0.0880)		(0.5294)
Master's	0.2242	0.2330	0.2245		0.9356	0.1679	0.1928*	0.1678		0.8527
	(0.1219)	(0.1224)	(0.1219)		(0.7154)	(0.0913)	(0.0918)	(0.0914)		(0.5464)
PhD	0.2589*	0.2636*	0.2595*		1.1708	0.2626*	0.2835**	0.2624*		1.6229**
	(0.1248)	(0.1254)	(0.1248)		(0.7319)	(0.1031)	(0.1035)	(0.1031)		(0.6068)
Same education	0.0412	0.0435	0.0413		0.2482	0.0174	0.0084	0.0173		0.0296
	(0.0239)	(0.0240)	(0.0239)		(0.1393)	(0.0307)	(0.0309)	(0.0307)		(0.1636)
Male more educated	0.0571	0.0646	0.0571		0.3556	-0.0057	-0.0098	-0.0057		-0.1400
	(0.0379)	(0.0381)	(0.0379)		(0.2166)	(0.0419)	(0.0422)	(0.0419)		(0.2352)
Non-rankable degree	0.2126	0.2371*	0.2140		0.8966	0.2125**	0.2201**	0.2123**		1.2286*
	(0.1143)	(0.1148)	(0.1143)		(0.6698)	(0.0822)	(0.0828)	(0.0823)		(0.4877)

Continued on next page

	Ads placed by females					Ads placed by males				
	Basic (1)	No caste (2)	Main caste (3)	Limited (4)	Logit (5)	Basic (6)	No caste (7)	Main caste (8)	Limited (9)	Logit (10)
Science	0.1002***	0.0951***	0.0999***		0.5945***	0.0456*	0.0423*	0.0457*		0.3074**
	(0.0214)	(0.0215)	(0.0214)		(0.1252)	(0.0192)	(0.0192)	(0.0192)		(0.1026)
Commerce	0.0529*	0.0525*	0.0526*		0.3096*	0.0781**	0.0819**	0.0785**		0.4895***
	(0.0222)	(0.0223)	(0.0222)		(0.1312)	(0.0259)	(0.0260)	(0.0259)		(0.1379)
Other field	0.0332	0 0321	0.0326		0.2229	0.0154	0.0162	0.0153		-0.2174
	(0.0518)	(0.0521)	(0.0518)		(0.2774)	(0.0742)	(0.0741)	(0.0742)		(0.4218)
Calcutta	0.0734***	0.0771***	0.0735***	0.0757***	0.4089***	0.0620**	0.0588**	0.0621**	0.0591**	0.3915***
	(0.0137)	(0.0138)	(0.0138)	(0.0138)	(0.0777)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.1064)
Same location	0.0469	0.0445	0.0463	0.0412	0.2988	-0.0437	-0.0455	-0.0438	-0.0442	-0.1492
	(0.0352)	(0.0353)	(0.0352)	(0.0352)	(0.2060)	(0.0289)	(0.0290)	(0.0289)	(0.0290)	(0.1593)
Same family origin	0.0348	0.0513**	0.0351	0.0363	0.2641*	0.0926***	0.1067***	0.0932***	0.0977***	0.6472***
	(0.0194)	(0.0194)	(0.0194)	(0.0194)	(0.1127)	(0.0214)	(0.0214)	(0.0214)	(0.0215)	(0.1246)
Log income	0.0995***	0.0953***	0.0992***		0.6010***					
	(0.0148)	(0.0148)	(0.0148)		(0.0853)					
Log wage	0.1046***	0.1093***	0.1050***		0.5581***					
	(0.0144)	(0.0145)	(0.0144)		(0.0837)					
Skin tone						-0.0506***	-0.0518***	-0.0508***	-0.0534***	-0.3004***
						(0.0101)	(0.0102)	(0.0101)	(0.0101)	(0.0595)
Beautiful						0.0071	0.0100	0.0071	0.0043	0.0920
						(0.0190)	(0.0191)	(0.0190)	(0.0191)	(0.1035)
Very beautiful						0.0532	0.0575	0.0533	0.0465	0.3279*
						(0.0300)	(0.0301)	(0.0300)	(0.0301)	(0.1569)
Predicted income				0.3478***					0.0817***	
				(0.0194)					(0.0229)	
N	5628	5628	5628	5628	5628	3944	3944	3944	3944	3944

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristic, age/height of the letter writer if no age/height was provided by the ad. age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Ads placed by females (males) received letters by males (females): the first five columns refer to decisions made by females regarding prospective grooms, the last five to decisions made by males regarding prospective brides. Standard errors in parentheses

* significant at 5%; ** significant at 1%; *** significant at 0 1%

Table 4: **Quality indices by caste categories**

	Ads placed by females			Ads placed by males		
	Own	Match	Share	Own	Match	Share
	(1)	(2)		(3)	(4)	
Panel A: By letters written by ad placers						
Any letter to caste above	0.0067	-0.0118	0.2558	-0.0360	-0.0122	0.3673
	(0.0147)	(0.0413)		(0.0365)	(0.0139)	
Any letter to caste below	-0.0072	-0.0526	0.3101	-0.0110	-0.0049	0.3673
	(0.0155)	(0.0382)		(0.0369)	(0.0207)	
N	123	37		41	23	
Panel B: By letters received by ad placers						
Any letter from caste above	-0.0101	0.0073	0.3981	0.0160	0.0255	0.5158
	(0.0066)	(0.0191)		(0.0111)	(0.0197)	
Any letter from caste below	0.0001	-0.0138*	0.5771	0.0163	0.0029	0.5860
	(0.0065)	(0.0066)		(0.0113)	(0.0067)	
N	285	158		526	131	

All cells correspond to a univariate regression of quality on a dummy variable indicating caste relationship. Standard errors in parentheses. Columns (1) and (3) refer to the quality of the ad-placer and columns (2) and (4) to the quality of the eventual match. Males (females) who place ads eventually marry females (males). Columns (2) and (3) are thus referring to quality of males while columns (1), (4) to quality of females.

* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table 5: **Dowries and probability of being considered**

	Full Regression		Parsimonious	
	Main effects in sample that does not mention dowries	Interaction of characteristics with no request for dowry	Main effects in sample that does not mention dowries	Interaction of characteristics with no request for dowry
	(1)	(2)	(3)	(4)
Same caste	0.0836**	0.1363	0.0887***	0.1971
	(0.0264)	(0.1080)	(0.0265)	(0.1070)
Diff. in caste*Higher caste male	0.0128	0.0089	0.0144	-0.0170
	(0.0143)	(0.0463)	(0.0144)	(0.0454)
Diff. in caste*Lower caste male	-0.0258*	0.0801	-0.0243	0.1018*
	(0.0124)	(0.0458)	(0.0124)	(0.0450)
Diff. in age	-0.0025	0.0031	-0.0040	0.0110
	(0.0049)	(0.0190)	(0.0049)	(0.0188)
Squared diff. in age	-0.0008**	-0.0001	-0.0008**	-0.0006
	(0.0003)	(0.0014)	(0.0003)	(0.0014)
Diff. in height	1.3842***	-1.9984	1.4127***	-2.1377*
	(0.2817)	(1.0405)	(0.2822)	(1.0249)
Squared diff. in height	-3.9449***	6.9149	-3.9571***	8.1506*
	(0.9871)	(3.6745)	(0.9880)	(3.5935)
High school	0.0776	-0.1167		
	(0.1100)	(0.1386)		
Post-secondary	0.1334	-0.2867		
	(0.1191)	(0.2939)		
Bachelor's	0.1239	-0.3886		
	(0.1187)	(0.2535)		
Master's	0.2513*	-0.4281		
	(0.1225)	(0.2641)		
PhD	0.2923*	-0.6111*		
	(0.1254)	(0.2697)		
Same education	0.0421	-0.3778		
	(0.0242)	(0.0638)		
Male more educated	0.0515	0.0639		
	(0.0383)	0.0882		
Non-rankable degree	0.2018			
	(0.1149)			
Science	0.0961***	0.0377		
	(0.0222)	(0.0809)		
Commerce	0.0467*	0.0654		
	(0.0232)	(0.0827)		
Other field	0.0232	0.0253		
	(0.0526)	(0.3418)		
Calcutta	0.0886***	0.1042*	0.0821***	-0.0916
	(0.0158)	(0.0482)	(0.0143)	(0.0520)
Same location	0.0792***	-0.0945	0.0442	0.0179
	(0.0143)	(0.0533)	(0.0358)	(0.0953)
Same family origin	0.0500	0.0535	0.0440*	-0.0142*
	(0.0358)	(0.0977)	(0.0199)	(0.0570)
Log income	0.0422*	-0.1274*		
	(0.0198)	(0.0583)		
Log wage	0.1084***	-0.0160		
	(0.0149)	(0.0565)		
Predicted income			0.3490***	0.0018
			(0.0198)	(0.0747)
No dowry	-0.3008		0.1042	
	(0.5804)		(0.7096)	
F-test: Same coefficients		1.24		1.34
N	5056		5056	

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Columns (1) and (2) represent the coefficients of a single regression. Columns (3) and (4) also represent a single regression. The main effects of each characteristics in the sample that does not mention dowries is presented in columns (1) and (3). The coefficients in columns (2) and (4) correspond to the coefficient of the interaction term between the letter stating that it has no dowry demand and each characteristic. Ads placed by females received letters by males - this table refers to decisions made by females regarding prospective grooms. Standard errors in parentheses

* significant at 5%; ** significant at 1%, *** significant at 0.1%

Table 6: **Difference in individuals' characteristics by marital status**

	Simulated			Observed	
	2.5 ptile (1)	97.5 ptile (2)	Mean (3)	2.5 ptile (4)	97.5 ptile (5)
	Panel A: Women, without search frictions				
Age	**0.8759**	**2.6992**	0.9215	0.2566	1.5865
Height	*-0.0246*	*-0.0063*	-0.0035	-0.0119	0.0049
Caste	*0.1842*	*1.0929*	-0.0772	-0.4235	0.2691
Education level	-1.0987	-0.6624	-0.1486	-0.3630	0.0658
Arts and Social Science	0.1242	0.3326	0.0148	-0.0899	0.1195
Commerce	*-0.1693*	*-0.0849*	-0.0416	-0.1118	0.0285
Science	*-0.2599*	*-0.0151*	0.0292	-0.0677	0.1260
Other field	**-0.0146**	**0.0318**	-0.0023	-0.0180	0.0133
From West Bengal	**-0.1472**	**0.0299**	0.0090	-0.1115	0.0935
Kolkota	*-0.5348*	*-0.1621*	-0.0290	-0.2126	0.1546
Skin rank	0.4877	0.8295	0.0214	-0.1407	0.1835
Very beautiful	**-0.0858**	**0.0059**	-0.0141	-0.0707	0.0425
Beautiful	**-0.2190**	**0.0428**	-0.0188	-0.1248	0.0873
Income	**-11265**	**3915**	-6267	-11449	-1084
Log wage	**-0.0770**	**0.0860**	0.0065	-0.1470	0.1599
"Quality"	-0.1134	-0.0838	-0.0050	-0.0088	0.0187
	Panel B: Women, with search frictions				
Age	**0.4462**	**2.1565**	0.9215	0.2566	1.5865
Height	*-0.0240*	*-0.0079*	-0.0035	-0.0119	0.0049
Caste	*0.1853*	*0.9895*	-0.0772	-0.4235	0.2691
Education level	-1.0220	-0.6292	-0.1486	-0.3630	0.0658
Arts and Social Science	0.1341	0.3701	0.0148	-0.0899	0.1195
Commerce	*-0.2080*	*-0.0937*	-0.0416	-0.1118	0.0285
Science	*-0.2660*	*-0.0049*	0.0292	-0.0677	0.1260
Other field	**-0.0190**	**0.0294**	-0.0023	-0.0180	0.0133
From West Bengal	**-0.1417**	**0.0363**	0.0090	-0.1115	0.0935
Kolkota	*-0.4092*	*-0.1001*	-0.0290	-0.2126	0.1546
Skin rank	0.4921	0.7767	0.0214	-0.1407	0.1835
Very beautiful	**-0.1042**	**0.0016**	-0.0141	-0.0707	0.0425
Beautiful	**-0.2086**	**0.0773**	-0.0188	-0.1248	0.0873
Income	*-1347*	*3853*	-6267	-11449	-1084
Log wage	**-0.1301**	**0.0820**	0.0065	-0.1470	0.1599
"Quality"	-0.1081	-0.0809	-0.0050	-0.0088	0.0187
	Panel C: Men, with search frictions				
Age	**-1.0919**	**0.5233**	0.4175	-0.6997	1.5346
Height	**-0.0179**	**0.0125**	-0.0040	-0.0206	0.0126
Caste	**-0.1533**	**2.0519**	0.1195	-0.3815	0.6205
Education level	*-1.2680*	*-0.5757*	-0.2399	-0.6066	0.1268
Arts and Social Science	**-0.0738**	**0.0811**	-0.0696	-0.1308	-0.0084
Commerce	**0.1040**	**0.4386**	0.1201	-0.0281	0.2683
Science	-0.5674	-0.2112	-0.0505	-0.2014	0.1004
Other field	**-0.0149**	**0.0224**	0.0000	0.0000	0.0000
Family origin	**-0.2584**	**0.1309**	0.0197	-0.1223	0.1617
Calcutta	**-0.5658**	**0.2069**	0.0363	-0.1122	0.1847
Income	*-8887*	*-2954*	-13560	-42033	14912
Log wage	-0.9925	-0.4129	-0.1141	-0.3196	0.0915
"Quality"	-0.1306	-0.0583	-0.0193	-0.0427	0.0041

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Table 7: **Couples' characteristics, simulated and observed**

	Simulated		Observed-considered			Observed-matched		
	2.5 ptile (1)	97.5 ptile (2)	Mean (3)	2.5 ptile (4)	97.5 ptile (5)	Mean (6)	2.5 ptile (7)	97.5 ptile (8)
Panel A: Without search frictions								
Age diff.	**5.3394**	**6.2323**	5.9032	5.8191	5.9873	5.6993	5.3476	6.0510
Age corr.	**0.7990**	**0.9242**	0.8331	0.8144	0.8507	0.6521	0.5700	0.7341
Height diff.	**0.1043**	**0.1235**	0.1201	0.1178	0.1223	0.1237	0.1146	0.1328
Height corr.	0.8108	0.9085	0.3825	0.3473	0.4188	0.3880	0.2875	0.4886
Same caste	0.8682	0.9732	0.7506	0.7333	0.7679	0.6937	0.6396	0.7478
Caste diff.	**0.0444**	**0.4856**	0.0916	0.0504	0.1328	-0.0071	-0.1584	0.1443
Caste corr.	**0.6536**	**0.9600**	0.8450	0.8202	0.8682	0.7599	0.6873	0.8324
Same education	**0.2529**	**0.7882**	0.4487	0.4299	0.4675	0.4380	0.3778	0.4982
Education diff.	-0.5093	0.0084	0.3385	0.3120	0.3823	0.2902	0.1393	0.4410
Education corr.	**0.2368**	**0.6001**	0.4202	0.3778	0.4620	0.3564	0.2383	0.4746
Same family origin	0.9898	1.0000	0.7839	0.7655	0.8024	0.7644	0.7060	0.8229
Family origin diff.	**-0.0047**	**0.0092**	0.0054	-0.0154	0.0263	0.0433	-0.0208	0.1073
Family origin corr.	0.9769	1.0000	0.5407	0.4959	0.5814	0.5147	0.3932	0.6361
Same residence	**0.0000**	**1.0000**	0.4687	0.4346	0.5028	0.4831	0.3834	0.5829
Location corr.	**-1.0000**	**0.4891**	0.0441	-0.0393	0.1195	-0.0566	-0.2246	0.2142
Log wage diff.	-0.4990	-0.0826	0.1375	0.0811	0.1939	0.2462	0.1349	0.3575
Log wage corr.	**-0.1670**	**0.4222**	0.0687	-0.0720	0.2017	0.1855	-0.1284	0.4993
Income diff.	**-11375**	**10300**	9277	-3842	22397	28374	-16	56764
Income corr.	**-0.6231**	**1.0000**	0.5760	0.4923	0.8139	0.4474	0.0837	0.8110
Quality diff.	*0.1299*	*0.1554*	0.1026	0.0983	0.1069	0.1202	0.1069	0.1336
Quality corr.	**0.0941**	**0.4640**	0.0386	-0.2434	0.3383	0.1950	0.0714	0.3187
Panel B: With search frictions								
Age diff.	**5.2017**	**6.2993**	5.9032	5.8191	5.9873	5.6993	5.3476	6.0510
Age corr.	**0.7700**	**0.9167**	0.8331	0.8144	0.8507	0.6521	0.5700	0.7341
Height diff.	**0.1036**	**0.1241**	0.1201	0.1178	0.1223	0.1237	0.1146	0.1328
Height corr.	0.7833	0.8920	0.3825	0.3473	0.4188	0.3880	0.2875	0.4886
Same caste	0.8869	0.9874	0.7506	0.7333	0.7679	0.6937	0.6396	0.7478
Caste diff.	**0.0040**	**0.4286**	0.0916	0.0504	0.1328	-0.0071	-0.1584	0.1443
Caste corr.	**0.6889**	**0.9915**	0.8450	0.8202	0.8682	0.7599	0.6873	0.8324
Same education	**0.2325**	**0.7870**	0.4487	0.4299	0.4675	0.4380	0.3778	0.4982
Education diff.	*-0.4397*	*0.1527*	0.3385	0.3120	0.3823	0.2902	0.1393	0.4410
Education corr.	**0.2223**	**0.6350**	0.4202	0.3778	0.4620	0.3564	0.2383	0.4746
Same family origin	0.9799	1.0000	0.7839	0.7655	0.8024	0.7644	0.7060	0.8229
Family origin diff.	**-0.0061**	**0.0149**	0.0054	-0.0154	0.0263	0.0433	-0.0208	0.1073
Family origin corr.	0.9524	1.0000	0.5407	0.4959	0.5814	0.5147	0.3932	0.6361
Same residence	**0.0000**	**1.0000**	0.4687	0.4346	0.5028	0.4831	0.3834	0.5829
Location corr.	**-0.7262**	**1.0000**	0.0441	-0.0393	0.1195	-0.0566	-0.2246	0.2142
Log wage diff.	-0.3845	0.0484	0.1375	0.0811	0.1939	0.2462	0.1349	0.3575
Log wage corr.	**-0.1770**	**0.4803**	0.0687	-0.0720	0.2017	0.1855	-0.1284	0.4993
Income diff.	**-6000**	**188000**	9277	-3842	22397	28374	-16	56764
Income corr.	**-1.0000**	**1.0000**	0.5760	0.4923	0.8139	0.4474	0.0837	0.8110
Quality diff.	*0.1310*	*0.1653*	0.1026	0.0983	0.1069	0.1202	0.1069	0.1336
Quality corr.	**0.0543**	**0.4191**	0.0386	-0.2434	0.3383	0.1950	0.0714	0.3187

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Table 8: **Couples' characteristics and the impact of caste, by caste**

	All castes		Brahmin		Kayastha		Baisya		Sagdope	
	2.5 ptile (1)	97.5 ptile (2)	2.5 ptile (3)	97.5 ptile (4)	2.5 ptile (5)	97.5 ptile (6)	2.5 ptile (7)	97.5 ptile (8)	2.5 ptile (9)	97.5 ptile (10)
Panel A: Without restrictions										
Age diff.	5.3394	6.2323	5.4830	6.3200	5.3668	6.1957	5.5092	6.2090	5.4749	6.1827
Age corr.	0.7990	0.9242	0.8677	0.9515	0.8697	0.9512	0.7453	0.8808	0.8018	0.9160
Height diff.	0.1043	0.1235	0.1086	0.1276	0.1035	0.1227	0.1057	0.1196	0.1065	0.1208
Height corr.	0.8108	0.9085	0.8590	0.9303	0.8466	0.9214	0.7170	0.8425	0.7740	0.8790
Same caste	0.8682	0.9732	0.7340	0.9899	0.9661	0.9991	0.9229	0.9946	0.7696	0.9790
Same education	0.2529	0.7882	0.2187	0.8429	0.2055	0.8016	0.3053	0.7483	0.2652	0.7877
Education diff.	-0.5093	0.0084	-0.5910	0.0262	-0.6129	-0.1270	-0.5431	-0.1430	-0.4906	-0.0257
Education corr.	0.2368	0.6001	0.3086	0.6688	0.2840	0.6453	0.2693	0.5692	0.2372	0.5628
Log wage diff.	-0.4990	-0.0826	-0.3596	-0.1905	-0.3894	-0.2215	-0.5133	-0.2609	-0.3747	-0 1432
Log wage corr.	-0.1670	0.4222	0.0651	0.2787	0.0120	0.2131	-0.0285	0.2019	-0.0442	0.2387
Quality diff.	0.1299	0.1554	0.1286	0.1512	0.1375	0.1513	0.1266	0.1488	0.1203	0.1452
Quality corr.	0.0941	0.4640	0.1419	0.4386	0.1034	0.3954	0.1456	0.3845	0.1365	0.3860
Panel B: With forced caste matching										
Age diff.	5.3814	6.2504	5.3744	6.5029	5.2848	6.2702	5.2521	6.4215	4.9047	6.2835
Age corr.	0.7856	0.9130	0.8176	0.9438	0.8413	0.9483	0.6697	0.8998	0.7200	0.9207
Height diff.	0.1050	0.1237	0.1050	0.1278	0.1033	0.1247	0.1012	0.1254	0.1039	0.1294
Height corr.	0.7998	0.8978	0.8624	0.9426	0.8350	0.9399	0.6714	0.8734	0.6927	0.9031
Same caste	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Same education	0.2612	0.7835	0.2034	0.8487	0.2127	0.8216	0.2959	0.7273	0.2143	0.8148
Education diff.	-0.4933	-0.0132	-0.6792	0.0508	-0.6028	0.0202	-0.5000	0.0556	-0.3333	0.4037
Education corr.	0.2538	0.6059	0.2106	0.7548	0.1849	0.6601	0.1375	0.5903	-0.1395	0.7290
Log wage diff.	-0.5338	-0.0920	-0.6701	0.0481	-0.7318	0.4171	-0.8300	-0.1611	-0.7702	0.3437
Log wage corr.	-0.1424	0.4106	-0.4029	0.4733	-0.8488	0.8865	-0.1616	0.9073	-0.9447	0.9537
Quality diff.	0.1297	0.1562	0.1218	0.1702	0.1118	0.1514	0.1286	0.1719	0.1040	0.1671
Quality corr.	0.0980	0.4547	0.0327	0.5188	0.0353	0.4921	0.0893	0.4734	-0.0952	0.5946
Panel C: Caste-blinded										
Age diff.	5.3867	6.2850	5.2343	6.2655	5.4810	6.4838	5.2844	6.3530	5.2500	6.3714
Age corr.	0.8818	0.9611	0.8382	0.9536	0.8706	0.9624	0.8910	0.9714	0.8947	0.9741
Height diff.	0.1039	0.1234	0.1031	0.1245	0.1037	0.1235	0.1004	0.1225	0.1026	0.1280
Height corr.	0.8937	0.9529	0.8887	0.9605	0.8849	0.9573	0.8900	0.9630	0.8797	0.9658
Same caste	0.1552	0.2357	0.1829	0.3690	0.2165	0.3904	0.0000	0.0862	0.0000	0.1622
Same education	0.2019	0.8503	0.2047	0.8731	0.2043	0.8507	0.2222	0.8969	0.1430	0.8846
Education diff.	-0.5890	0.0268	-0.6240	0.0842	-0.6621	0.0299	-0.5911	0.1031	-0.5963	0.3513
Education corr.	0.2913	0.6902	0.2479	0.7807	0.2161	0.7153	0.2584	0.7994	-0.0391	0.7909
Log wage diff.	-0.4723	-0.0717	-0.6604	0.0217	-0.6750	0.3825	-0.7236	-0.0225	-0.6789	0.4324
Log wage corr.	-0.1366	0.4105	-0.3681	0.5017	-0.6788	0.8421	-0.2646	0.7928	-0.8874	0.8542
Quality diff.	0.1284	0.1562	0.1315	0.1780	0.1091	0.1529	0.1304	0.1775	0.0834	0.1501
Quality corr.	0.0888	0.5048	0.0301	0.5254	0.0588	0.5425	0.0929	0.5813	-0.0936	0.6616

Table 9: **Distribution of costs of...**

	Keeping caste		**Education**	
	Male	Female	Male	Female
Education	-0.0757	0.0373		
	[-0.8269, 0.6953]	[-0.5581, 0.6194]		
Height difference	-0.0001	0.0083	0.1488	2.7930
	[-0.0182, 0.0181]	[-0.0118, 0.0303]	[-5.2043, 5.2367]	[-1.9975, 7.2504]
Age difference	0.2053	-0.1221	-0.0667	**-0.1878**
	[-0.8727, 1.5003]	[-1.2574, 1.0445]	[-0.1803, 0.0396]	**[-0.25566, -0.1158]**
Income	-2628.65	7357.71	-0.0025	
	[-67460.78, 47564.81]	[-927.18, 13233.08]	[-0.0181, 0.0108]	
Wage	-0.1232	0.0836	0.2847	
	[-0.5854, 0.3512]	[-0.6631, 0.8667]	[-0.0602, 0.6329]	
Very beautiful		-0.0134		**-0.3645**
		[-0.2929, 0.1769]		**[-0.5998, -0.1351]**
Beautiful		0.0671		-0.1266
		[-0.3631, 0.4618]		[-0.3220, 0.0495]
Skin tone		-0.0684		0.1472
		[-0.7873, 0.5495]		[-0.0042, 0.3224]

The 2.5 and 97.5 percentile of the distribution of coefficients is presented in brackets. Bold entries mark significance at 5% or more.

Figure 1: **Indifference curve of males**

Figure 2: **Indifference curve of females**

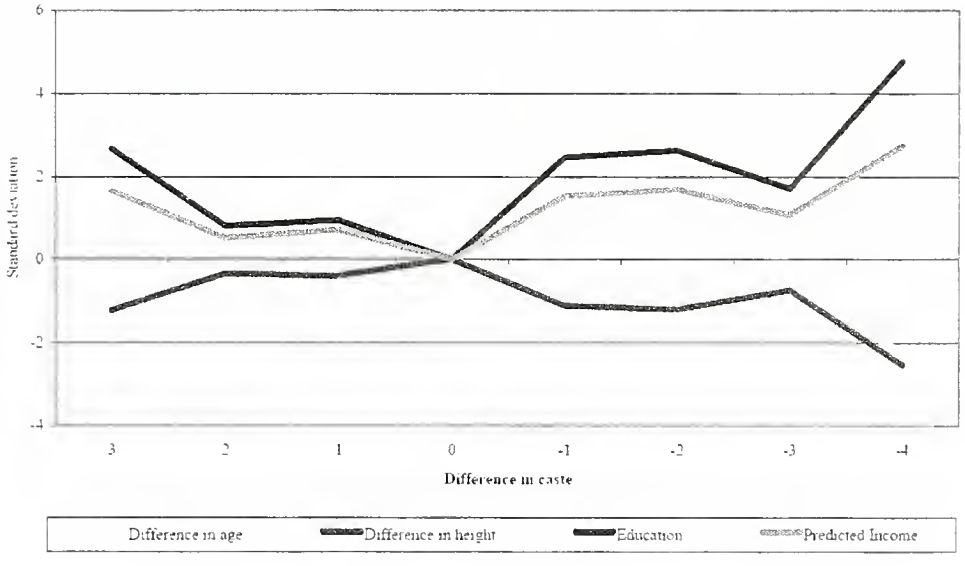


Figure 3: **Proportion of considered letters by quality of the letter and ad placer, ads placed by females**

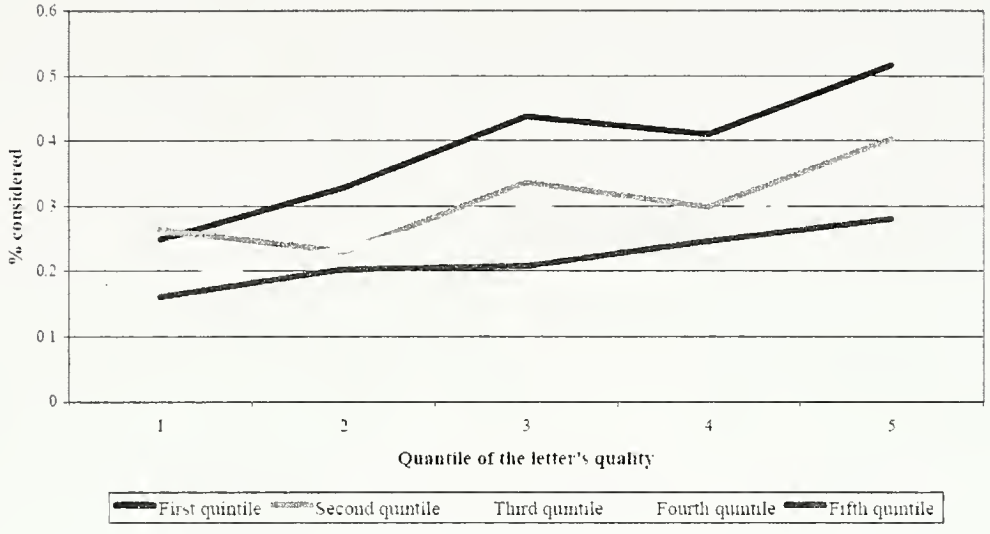


Figure 4: **Proportion of considered letters by quality of the letter and ad placer, ads placed by males**

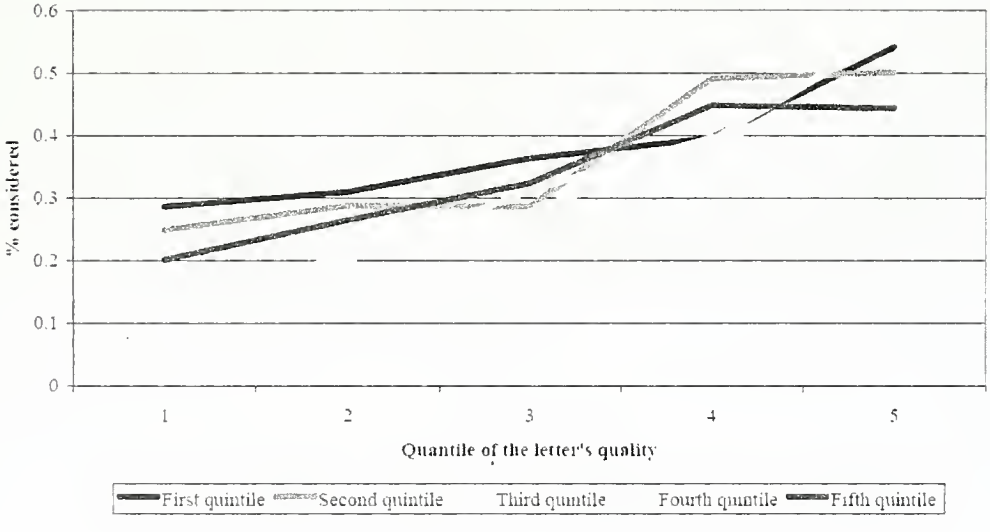
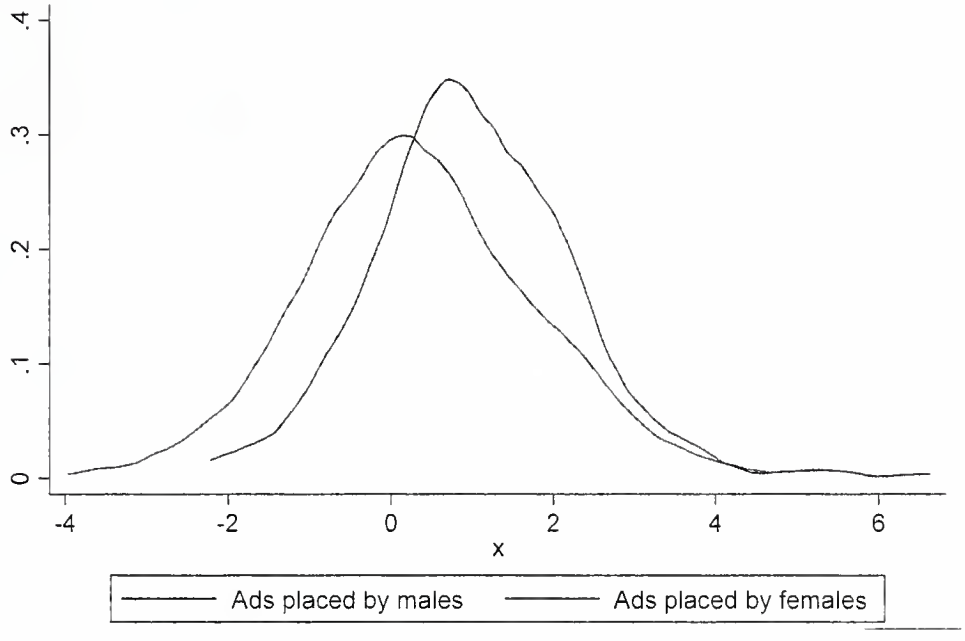


Figure 5: **Distribution of preferences for own caste**



A Theoretical Appendix: Adding unobserved characteristics

This section proves that if exploration is not too costly, what individuals choose to be the set of options they explore reflects their true ordering over observables, even in the presence of an unobservable characteristic they may also care about.

Formally, we assume that in addition to the two characteristics already in our model, x and y, there is another (payoff-relevant) characteristic z (such as demand for dowry) not observed by the respondent that may be correlated with x. Is it a problem for our empirical analysis that the decision-maker can make inferences about z from their observation of x? The short answer, which this section briefly explains, is no, as long as the cost of exploration (upon which z is revealed) is low enough.

Suppose $z \in \{H, L\}$ with $H > L$ (say, the man is attractive or not). Let us modify the payoff of a woman of caste j and type y who is matched with a man of caste i and type (x, z) to $u^W(i, j, x, y) = A(j, i)f(x, y)z$. Let the conditional probability of z upon observing x, is denoted by $p(z|x)$. Given z is binary, $p(H|x) + p(L|x) = 1$. In that case, the expected payoff of this woman is:

$$A(j, i)f(x, y)p(H|x)H + A(j, i)f(x, y)p(L|x)L.$$

Suppose the choice is between two men of caste i whose characteristics are x' and x'' with $x'' > x'$. If x and z are independent (i.e., $p(z|x) = p(z)$ for $z = H, L$ for all x), or, x and z are positively correlated, then clearly the choice will be x''. Similarly, if it is costless to contact someone with type x'' and find out about z (both in terms of any direct cost, as well as indirect cost of losing out on the option x') the choice, once again, will be x'' independent of how (negatively) correlated x and z are.

More formally, for this simple case, suppose we allow x and z to be correlated in the following way: $p(H|x'') = p\mu, p(L|x'') = 1 - p\mu$, $p(H|x') = p$, and $p(L|x') = 1 - p$. If $\mu > 1$ we have positive correlation between z and x, if $\mu < 1$ we have negative correlation, and if $\mu = 1$, x and z are independent. Suppose exploring a single option costs c. Let us assume that $Hf(x', y) > Lf(x'', y)$ – otherwise, it is a dominant strategy to explore x'' only.

We consider two strategies. One is to explore only one of the two options and stick with the choice independent of the realization of z. The other is to explore both the options at first, and discard one of them later.

If the decision-maker explores both options, the choice will be x'' if either the z associated with it is H or if both x'' and x' have $z = L$ associated with them. Otherwise, the choice will be x'. The *ex ante* expected payoff from this strategy is

$$p\mu Hf(x'', y) + (1 - p\mu)[(1 - p)Lf(x'', y) + pHf(x', y)] - 2c.$$

This is obviously more than what he gets by exploring either one alone (namely, $f(x', y)\{pH + (1-p)L\} - c$ or $f(x'', y)\{p\mu H + (1-p\mu)L\} - c$) as long as c is small enough for any fixed value of $\mu > 0$.

Proposition 4 *For any fixed value of $\mu > 0$, so long as the exploration cost c is small enough, x'' will be chosen at the exploration stage whenever x' is chosen.*

In other words, as long as exploration is not too costly, what people choose to be the set of options to explore reflects their true ordering over the observables. In other words the indifference curve we infer from the "up or out" choices reflects their true preferences over the set of observables.

B Data Appendix

Ads and letters provided very rich qualitative information that had to be coded to make the data analysis possible. We first coded caste, using the process described in the text.

Second, we coded information provided on education levels. Educational attainment was classified into seven categories: less than high school, high school completion, non-university post-secondary, bachelor's, master's, PhD or professional degree and non-classifiable degree.[36] In addition, we also coded, when available, the field in which the degree was obtained. We sorted these into four groups: humanities and social sciences (B.A, B.Ed, M.A, etc.), commerce (B.Com, MBA), science (B.Sc., B.Eng, M.Sc., etc.) and other fields (law, religion, etc.).

Third, we coded the available information on earning levels. When provided in the ad, self-reported earnings were converted into a monthly figure. This value will be referred to as "income." In addition, when the ad-placer or the letter writer provided his or her occupation, we used the National Sample Survey of India to construct an occupational score for the occupation (we refer to this below as "wage"). Note that prospective brides almost never report this information, and it will therefore be used only for the letters and ads from prospective grooms.

Fourth, we coded information on the origin of the family (East or West Bengal) and the current location of the prospective bride or groom under the following categories: Kolkata, Mumbai, other West Bengal, or other (mainly, abroad).[37]

Fifth, a very large fraction of ads from prospective brides specify physical characteristics of the women, using fairly uniform language and the same broad characteristics. Skin color was

[36] This last group mostly includes degrees in computer science from private institutions that were difficult to place within the existing ranking.

[37] At the time of Independence, the state of Bengal was partitioned into two states, one that remained in India, West Bengal, and the other that joined Pakistan, East Pakistan (which later became Bangladesh) . Many Hindus migrated from East to West Bengal. There are some variations in terms of dialect, cultural and social norms among Bengalis depending on their family origin. This has some relevance in the arranged marriage market.

coded into four categories (from "extremely fair" to "dark") and we associate each category with a number from 1 to 4, with higher numbers representing darker skins. General beauty was divided into three categories ("very beautiful," "beautiful" and "decent-looking").

Finally, ads occasionally mention a multitude of other characteristics, such as "gotras" (a sub-group within one's caste based on lineage such that inter-marriages are ruled out under exogamy), astrological signs, blood type, family characteristics, personality traits, previous marital history, and specific demands. These were coded as well. However, each of these is rarely mentioned and so including or excluding them does not affect our results.

C Appendix tables

Table C.1: **Characteristics of ads by attrition status in second round interviews**

Variable	Ads placed by females				Ads placed by males			
	Means		Difference		Means		Difference	
	Found	Not found	Mean	Sd. Error	Found	Not found	Mean	Sd. Error
Number of responses	23.004	18.000	5.00	4.65	79.874	89.071	-9.20	19.88
Caste								
Brahmin	0.27	0.21	0.06	0.10	0.25	0.29	-0.03	0.12
Baidya	0.04	0.16	**-0.12**	**0.05**	0.05	0.00	0.05	0.06
Kshatriya	0.02	0.00	0.02	0.03	0.02	0.00	0.02	0.03
Kayastha	0.35	0.21	0.14	0.11	0.31	0.36	-0.04	0.13
Baisya and others	0.19	0.21	-0.03	0.09	0.18	0.14	0.04	0.11
Sagdope and others	0.10	0.16	-0.06	0.07	0.12	0.14	-0.02	0.09
Other castes	0.02	0.00	0.02	0.03	0.02	0.07	-0.05	0.04
Scheduled castes	0.02	0.05	-0.03	0.04	0.05	0.00	0.05	0.06
Physical characteristics								
Age	26.55	27.67	-1.12	0.88	32.17	31.50	0.67	1.32
Height (meters)	1.58	1.59	-0.01	0.01	1.70	1.68	0.03	0.02
Skin tone	2.30	2.36	-0.06	0.22				
Very beautiful	0.08	0.20	*-0.12*	*0.07*				
Beautiful	0.44	0.53	-0.09	0.13				
Education and Income								
Less than high school	0.02	0.06	-0.03	0.04	0.01	0.00	0.01	0.03
High school	0.09	0.06	0.04	0.07	0.10	0.00	0.10	0.08
Post-secondary	0.00	0.00	0.00	0.01	0.06	0.00	0.06	0.06
College	0.53	0.50	0.03	0.12	0.42	0.46	-0.04	0.14
Master's	0.28	0.33	-0.05	0.11	0.18	0.23	-0.05	0.11
PhD	0.06	0.06	0.00	0.06	0.22	0.31	-0.09	0.12
Other degree	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.03
Humanities/Arts	0.57	0.75	-0.18	0.13	0.04	0.09	-0.05	0.07
Commerce	0.13	0.06	0.06	0.08	0.41	0.27	0.14	0.15
Science	0.30	0.19	0.11	0.12	0.55	0.64	-0.09	0.16
Other field	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Log wage	5.56	5.41	0.15	0.14	5.61	5.61	0.00	0.21
Log income	8.68	9.16	-0.48	0.60	9.45	9.22	0.23	0.39
Location								
Calcutta	0.82	0.60	0.22	0.18	0.78	0.40	*0.38*	*0.19*
West Bengali	0.39	0.40	-0.01	0.13	0.38	0.56	-0.17	0.17
Demands mentioned								
Only within caste	0.10	0.05	0.05	0.07	0.09	0.07	0.02	0.08
Caste no bar	0.32	0.42	-0.10	0.11	0.24	0.29	-0.05	0.12
No dowry demanded	0.01	0.05	-0.04	0.03	0.10	0.14	-0.04	0.08
Ads which omit...								
Caste	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.02
Age	0.01	0.05	-0.04	0.03	0.03	0.14	**-0.11**	**0.05**
Height	0.03	0.11	-0.07	0.04	0.11	0.14	-0.04	0.09
Education	0.08	0.05	0.03	0.06	0.19	0.07	0.12	0.11
Field	0.25	0.16	0.10	0.10	0.30	0.21	0.09	0.13
Residence	0.84	0.74	0.11	0.09	0.51	0.64	-0.13	0.14
Family origin	0.23	0.21	0.02	0.10	0.28	0.36	-0.08	0.12
Wage	0.85	0.63	**0.22**	**0.09**	0.57	0.50	0.07	0.14
Income	0.98	0.89	**0.08**	**0.04**	0.73	0.79	-0.05	0.12
Skin tone	0.21	0.26	-0.06	0.10				
Beauty	0.27	0.21	0.06	0.10				

Differences in italics are significant at 10%, those in bold, at 5%.

Table C.2: **Characteristics of ads who agreed and refused second round interview**

Variable	Ads placed by females				Ads placed by males			
	Means		Difference		Means		Difference	
	Agreed	Refused	Mean	Sd. Error	Agreed	Refused	Mean	Sd. Error
Number of responses	25.643	18.844	*6.80*	*3.51*	85.551	71.217	14.33	17.17
Caste								
Brahmin	0.25	0.25	0.00	0.08	0.23	0.36	-0.13	0.09
Baidya	0.04	0.06	-0.02	0.04	0.06	0.08	-0.02	0.05
Kshatriya	0.03	0.00	0.03	0.03	0.03	0.00	0.03	0.03
Kayastha	0.39	0.31	0.08	0.09	0.28	0.28	0.00	0.10
Baisya and others	0.18	0.16	0.03	0.07	0.21	0.12	0.09	0.09
Sagdope and others	0.07	0.16	-0.09	0.05	0.13	0.04	0.09	0.07
Other castes	0.02	0.03	-0.01	0.03	0.03	0.00	0.03	0.03
Scheduled castes	0.03	0.03	-0.01	0.03	0.02	0.12	**-0.10**	**0.04**
Physical characteristics								
Age	25.88	26.53	-0.65	0.60	31.92	32.45	-0.53	0.98
Height (meters)	1.58	1.59	-0.01	0.01	1.71	1.70	0.01	0.02
Skin tone	2.30	2.23	0.07	0.16				
Very beautiful	0.10	0.00	*0.10*	*0.06*				
Beautiful	0.42	0.58	-0.15	0.11				
Education and Income								
Less than high school	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.02
High school	0.10	0.03	0.06	0.06	0.10	0.05	0.05	0.07
Post-secondary	0.01	0.00	0.01	0.02	0.04	0.05	-0.01	0.05
College	0.51	0.53	-0.02	0.10	0.42	0.37	0.05	0.12
Master's	0.29	0.37	-0.08	0.09	0.22	0.16	0.07	0.10
PhD	0.07	0.07	0.00	0.05	0.20	0.37	*-0.17*	*0.10*
Other degree	0.01	0.00	0.01	0.02	0.01	0.00	0.01	0.02
Humanities/Arts	0.59	0.42	0.17	0.11	0.07	0.06	0.02	0.07
Commerce	0.13	0.27	*-0.14*	*0.08*	0.38	0.28	0.10	0.12
Science	0.28	0.31	-0.03	0.10	0.55	0.67	-0.12	0.13
Other field	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00
Log wage	5.53	5.73	*-0.21*	*0.12*	5.66	5.57	0.09	0.15
Log income	9.39	8.52	*0.87*	*0.28*	9.52	9.49	0.04	0.33
Location								
Calcutta	0.88	0.60	0.28	0.18	0.78	0.64	0.14	0.14
West Bengali	0.42	0.30	0.11	0.11	0.40	0.26	0.13	0.12
Demands mentioned								
Only within caste	0.09	0.09	0.00	0.06	0.08	0.04	0.04	0.06
Caste no bar	0.34	0.31	0.02	0.09	0.27	0.08	**0.19**	**0.09**
No dowry demanded	0.02	0.00	0.02	0.02	0.10	0.08	0.02	0.06
Ads which omit...								
Caste	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.02
Age	0.01	0.00	0.01	0.01	0.02	0.12	**-0.10**	**0.04**
Height	0.03	0.00	0.03	0.03	0.11	0.20	-0.09	0.07
Education	0.08	0.06	0.01	0.05	0.15	0.24	-0.09	0.08
Field	0.25	0.19	0.06	0.08	0.26	0.28	-0.02	0.10
Residence	0.84	0.84	0.00	0.07	0.51	0.56	-0.05	0.11
Family origin	0.24	0.28	-0.04	0.08	0.31	0.24	0.07	0.10
Wage	0.83	0.88	-0.05	0.07	0.54	0.44	0.10	0.11
Income	0.97	0.97	0.01	0.03	0.74	0.72	0.02	0.10
Skin tone	0.22	0.06	**0.16**	**0.08**				
Beauty	0.27	0.19	0.08	0.08				

Differences in italics are significant at 10%, those in bold, at 5%

Table C.3: **Caste groupings**

	1. Brahmin	
Brahmin	Kshatriya Brahmin	Rudraja Brahmin*
Kulin Brahmin	Nath Brahmin	Baishnab Brahmin*
Sabitri Brahmin	Rajput Brahmin	Baishnab*
Debnath Brahmin	Gouriya Baishnab*	Nath*
Kanya Kubja Brahmin		
	2. Baidya	
Baidya	Lata Baidya	Kulin Baidya
Rajasree Baidya		
	3. Kshatriya	
Kshatriya	Ugra Kshatriya	Rajput (Solanki) Kshatriya
Poundra Kshatriya	Malla Kshatriya	Jana Kshatriya
Rajput Kshatriya	Barga Kshatriya	
	4. Kayastha	
Kayastha	Rajput Kayastha	Kayastha Karmakar
Kulin Kayastha	Pura Kayastha	Karmakar
Kshatriya Kayastha	Mitra Mustafi	Mitra Barujibi
Kshatriya Karmakar		
	5. Baisya and others	
Baisya	Suri	Teli
Baisya Saha	Suri Saha	Ekadash Teli
Baisya Ray	Rudra Paul	Dadash Teli
Baisya Kapali	Modak	Tili
Baisya Teli	Modak Moyra	Ekadash Tili
Rajasthani Baisya	Banik	Dsadah Tili
Barujibi	Gandha Banik	Marwari
Baisya Barujibi	Kangsha Banik	Malakar
Sutradhar	Khandagrami Subarna Banik	Tambuli
Baisya Sutradhar	Subarna Banik	Rajak
Tantubai	Shankha Banik	Kasari
Baisya Tantubai	Swarnakar	Baisya Tambuli
	6. Sadgope and others	
Sadgope	Yadav	Mahishya
Kulin Sadgope	Yadav Ghosh	Kumbhakar
Kshatriya Sadgope	Goyala	Satchasi
Yadav (Gope)	Gope	
	7. Other (mostly) non-scheduled castes	
Kaibarta	Rajak	Paramanik
Jele	Bauri	Jelia Kaibarta
Napit		
	8. (mostly) Scheduled castes	
Rajbanshi	Namasudra	Karan
Rajbanshi Kshatriya	Sagari	SC
Malo	Sudra	OBC
Mathra	Baisya Rajbanshi	

Table C.4: **Rank of the letter**

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Same caste	1.2797***		1.1275**	1.3320***	0.4314***	1.2591***		1.5022***	1.4068***	0.3595***
	(0.2933)		(0.3821)	(0.4171)	(0.0928)	(0.3458)		(0.4292)	(0.4153)	(0.0928)
Same main caste			0.2377					-0.4295		
			(0.3825)					(0.4490)		
Diff. in caste*Higher caste male	-0.0500		-0.0179	-0.0176	-0.0034	-0.4707**		-0.5472**	-0.3732	-0.1421**
	(0.1341)		(0.1437)	(0.1615)	(0.0418)	(0.1699)		(0.1878)	(0.2379)	(0.0461)
Diff. in caste*Lower caste male	0.1070		0.0767	0.0781	0.0281	-0.3310		-0.2548	-0.3634	-0.0976*
	(0.1183)		(0.1280)	(0.1742)	(0.0372)	(0.1705)		(0.1882)	(0.2093)	(0.0458)
Same caste*only within	1.1726		1.1737	1.1669	0.2128	2.1112		2.0985	2.1664	0.7029
	(0.9116)		(0.9117)	(1.0856)	(0.2848)	(1.3256)		(1.3257)	(1.4721)	(0.3674)
Diff. in caste*only within	-0.4459		-0.4471	-0.4554	-0.1670	0.0183		0.0094	-0.1336	0.0874
	(0.3334)		(0.3334)	(0.4955)	(0.1117)	(0.5781)		(0.5782)	(0.6173)	(0.1582)
Same caste*no bar	-0.8681**		-0.8678**	-0.8606*	-0.2911**	-0.8599*		-0.8912*	-0.9364	-0.2521*
	(0.3258)		(0.3258)	(0.3821)	(0.1028)	(0.4315)		(0.4328)	(0.5233)	(0.1156)
Diff. in caste*no bar	-0.1021		-0.1041	-0.0832	-0.0247	0.2092		0.2020	0.1752	0.0734
	(0.1071)		(0.1072)	(0.1240)	(0.0342)	(0.1521)		(0.1523)	(0.2094)	(0.0409)
Diff. in age	0.0345	0.0255	0.0348	0.0217	0.0053	0.5215***	0.5411***	0.5205***	0.4459*	0.1457***
	(0.0405)	(0.0405)	(0.0405)	(0.0856)	(0.0127)	(0.0816)	(0.0820)	(0.0816)	(0.1978)	(0.0218)
Squared diff. in age	-0.0114***	-0.0115***	-0.0114***	-0.0110*	-0.0031***	-0.0284***	-0.0291***	-0.0282***	-0.0263*	-0.0079***
	(0.0023)	(0.0023)	(0.0023)	(0.0054)	(0.0007)	(0.0057)	(0.0057)	(0.0057)	(0.0117)	(0.0015)
Diff. in height	9.5137***	9.8711***	9.4794***	9.8330**	3.5492***	7.2790*	6.8472*	7.2231*	7.6700*	1.9194*
	(2.5694)	(2.5757)	(2.5701)	(3.3196)	(0.8651)	(3.2304)	(3.2517)	(3.2309)	(3.5366)	(0.8796)
Squared diff. in height	-24.5037**	-26.3139**	-24.4011**	-25.3370	-9.5136**	-69.0103***	-68.9625***	-68.8785***	-70.3327***	-18.7289***
	(9.2415)	(9.2562)	(9.2436)	(13.7834)	(3.2019)	(12.3135)	(12.3931)	(12.3145)	(13.6789)	(3.3576)
High school	0.6719	0.9189	0.6811		0.3796	1.7107**	1.7634**	1.7049**		0.4798**
	(0.9403)	(0.9438)	(0.9405)		(0.3366)	(0.6092)	(0.6140)	(0.6092)		(0.1709)
Post-secondary	1.3963	1.7144	1.4059		0.5588	2.5003	2.3729	2.4921		0.6638
	(1.0262)	(1.0290)	(1.0264)		(0.3629)	(1.4645)	(1.4709)	(1.4645)		(0.3922)
Bachelor's	1.4920	1.7376	1.4965		0.6384	2.7817**	2.9152**	2.7961**		0.7474**
	(1.0213)	(1.0243)	(1.0214)		(0.3635)	(0.8894)	(0.8959)	(0.8896)		(0.2434)
Master's	2.3654*	2.6088*	2.3650*		0.9383*	3.9425***	4.0203***	3.9590***		1.0457***
	(1.0533)	(1.0564)	(1.0534)		(0.3739)	(0.9236)	(0.9303)	(0.9237)		(0.2527)

Continued on next page

	Ads placed by females					Ads placed by males				
	Basic	No caste	Main caste	Limited	Logit	Basic	No caste	Main caste	Limited	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PhD	2.6963*	2.9129**	2.6967*		1.0487**	4.2363***	4.2562***	4.2333***		1.2354***
	(1.0810)	(1.0842)	(1.0811)		(0.3828)	(1.0650)	(1.0720)	(1.0650)		(0.2918)
Same education	0.5329*	0.5361*	0.5340*		0.1369*	0.2423	0.1380	0.2433		0.0577
	(0.2091)	(0.2100)	(0.2092)		(0.0662)	(0.2995)	(0.3013)	(0.2995)		(0.0803)
Male more educated	0.8218*	0.8550*	0.8256*		0.2317*	0.3416	0.2331	0.3442		0.0886
	(0.3315)	(0.3327)	(0.3316)		(0.1065)	(0.4169)	(0.4194)	(0.4169)		(0.1120)
Non-rankable degree	1.8538	2.1751*	1.8618		0.7512*	2.6315**	2.6192**	2.6275**		0.7227**
	(0.9855)	(0.9886)	(0.9857)		(0.3497)	(0.8065)	(0.8122)	(0.8065)		(0.2225)
Science	1.0444***	0.9810***	1.0454***		0.3522***	0.7039***	0.6512***	0.7092***		0.2050***
	(0.1882)	(0.1887)	(0.1882)		(0.0600)	(0.1928)	(0.1931)	(0.1929)		(0.0516)
Commerce	0.3640	0.3573	0.3646		0.1096	1.1107***	1.1203***	1.1076***		0.3257***
	(0.1948)	(0.1956)	(0.1948)		(0.0622)	(0.2600)	(0.2612)	(0.2600)		(0.0698)
Other field	0.1361	0.1378	0.1388		0.0921	1.1653	1.2332	1.1686		0.3351
	(0.4631)	(0.4654)	(0.4632)		(0.1476)	(0.7950)	(0.7994)	(0.7950)		(0.2213)
Calcutta	0.4690***	0.4953***	0.4703***	0.4923**	0.1738***	0.6515***	0.6240**	0.6501***	0.6288**	0.1741***
	(0.1204)	(0.1206)	(0.1205)	(0.1744)	(0.0383)	(0.1891)	(0.1897)	(0.1891)	(0.2223)	(0.0509)
Same location	0.4846	0.4160	0.4831	0.4059	0.1181	-0.1912	-0.2096	-0.1944	-0.2074	-0.0551
	(0.3086)	(0.3097)	(0.3086)	(0.3802)	(0.0959)	(0.2876)	(0.2893)	(0.2877)	(0.3468)	(0.0777)
Same family origin	0.2665	0.3861*	0.2770	0.2767	0.0712	0.7190***	0.8573***	0.7150***	0.8038**	0.1903**
	(0.1710)	(0.1710)	(0.1710)	(0.1963)	(0.0538)	(0.2156)	(0.2163)	(0.2156)	(0.2646)	(0.0580)
Log income	0.8761***	0.8254***	0.8782***		0.2906***					
	(0.1310)	(0.1308)	(0.1310)		(0.0431)					
Log wage	0.9205***	0.9451***	0.9221***		0.2988***					
	(0.1258)	(0.1262)	(0.1259)		(0.0397)					
Skin tone						-0.4585***	-0.4657***	-0.4581***	-0.4994***	-0.1292***
						(0.1005)	(0.1012)	(0.1005)	(0.1531)	(0.0271)
Beautiful						0.2045	0.2127	0.2095	0.1766	0.0404
						(0.1885)	(0.1893)	(0.1885)	(0.2272)	(0.0505)
Very beautiful						0.5376	0.5587	0.5363	0.4247	0.1614*
						(0.2934)	(0.2951)	(0.2934)	(0.3056)	(0.0787)
Predicted income				3.2509*					0.8807	
				(1.3526)					(2.2178)	
N	5094	5094	5094	5094	5094	3520	3520	3520	3520	3520

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Ads placed by females (males) received letters by males (females); the first five columns refer to decisions made by females regarding prospective grooms, the last five to decisions made by males regarding prospective brides. Standard errors in parentheses.

* significant at 5%; ** significant at 1%, *** significant at 0.1%

Table C.5: **Responses to "not too good" letters**

	Ads placed by females				Ads placed by males			
	Considered		Rank		Considered		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same caste	0.1073*	0.1134**	1.0817*	1.2763***	0.0884	0.1498***	1.2144*	1.4484***
	(0.0451)	(0.0364)	(0.4438)	(0.3404)	(0.0489)	(0.0418)	(0.5085)	(0.4270)
Diff. in caste*	0.0464*	0.0253	0.2376	0.0389	0.0570*	0.0186	0.7497**	0.4847*
Higher caste male	(0.0197)	(0.0166)	(0.1888)	(0.1524)	(0.0243)	(0.0203)	(0.2536)	(0.2100)
Diff. in caste*	0.0027	-0.0058	-0.0291	-0.1165	0.0373	0.0431*	0.6135*	0.5529**
Lower caste male	(0.0175)	(0.0146)	(0.1714)	(0.1356)	(0.0233)	(0.0200)	(0.2464)	(0.2060)
Same caste*only within	-0.0906	-0.0344	-1.0448	-0.6513	0.1245	0.1138	0.4840	0.6478
	(0.1408)	(0.1273)	(1.2780)	(1.1149)	(0.1851)	(0.1679)	(1.8022)	(1.6123)
Diff. in caste*only within	0.0036	0.0062	0.3854	0.5496	0.0096	0.0088	0.5102	0.6311
	(0.0492)	(0.0473)	(0.4439)	(0.4123)	(0.0797)	(0.0751)	(0.7765)	(0.7210)
Same caste*no bar	-0.0733	-0.0527	-0.9739*	-1.0054**	0.0027	-0.0206	-0.4229	-0.9570
	(0.0508)	(0.0415)	(0.4908)	(0.3853)	(0.0574)	(0.0499)	(0.6295)	(0.5286)
Diff. in caste*no bar	0.0031	0.0069	0.1017	0.1457	-0.0265	-0.0066	-0.5458*	-0.3208
	(0.0163)	(0.0135)	(0.1559)	(0.1243)	(0.0206)	(0.0175)	(0.2236)	(0.1847)
Diff. in age	0.0058	0.0053	0.0372	0.0696	0.0435***	0.0436***	0.5121***	0.4841***
	(0.0060)	(0.0051)	(0.0560)	(0.0459)	(0.0120)	(0.0105)	(0.1297)	(0.1103)
Squared diff. in age	-0.0008*	-0.0009**	-0.0097***	-0.0117***	-0.0023*	-0.0021**	-0.0270*	-0.0245**
	(0.0003)	(0.0003)	(0.0028)	(0.0025)	(0.0009)	(0.0008)	(0.0105)	(0.0085)
Diff. in height	0.9198*	0.7934*	9.2645*	6.8037*	0.7503	0.9038*	6.2082	7.4802*
	(0.4189)	(0.3334)	(4.1113)	(3.2594)	(0.4284)	(0.3641)	(4.3149)	(3.5929)
Squared diff. in height	-3.2350	-2.0427	-25.9230	-13.3929	-6.1195***	-6.0644***	-66.2058***	-65.7108***
	(1.7081)	(1.2791)	(16.7790)	(12.7629)	(1.4949)	(1.3248)	(15.1818)	(13.3146)
High school	-0.0930	-0.0507	-0.0679	0.3281	0.1697	0.1437	2.9543*	2.0051**
	(0.2237)	(0.1441)	(2.0167)	(1.2549)	(0.1245)	(0.0766)	(1.2073)	(0.7601)
Post-secondary	0.0173	0.0473	1.0474	1.2573	0.3295	0.2195	4.5315*	2.4932
	(0.2323)	(0.1522)	(2.1097)	(1.3380)	(0.2200)	(0.1573)	(2.2618)	(1.6627)
Bachelor's	-0.0341	0.0017	1.3182	1.2914	0.1965	0.1959	4.4956**	2.9271**
	(0.2323)	(0.1523)	(2.1078)	(1.3368)	(0.1488)	(0.1041)	(1.4671)	(1.0621)
Master's	0.0745	0.1415	2.1164	2.3877	0.3004*	0.2742*	5.8510***	4.1727***
	(0.2374)	(0.1559)	(2.1598)	(1.3715)	(0.1530)	(0.1080)	(1.5109)	(1.1016)
PhD	0.1705	0.1858	3.2869	2.9018*	0.3640	0.3425**	6.2600**	5.9120***
	(0.2413)	(0.1597)	(2.1997)	(1.4062)	(0.1920)	(0.1321)	(1.9928)	(1.4177)
Same education	0.0579	0.0432	0.3489	0.5761*	-0.0065	0.0194	0.1562	0.3351
	(0.0342)	(0.0273)	(0.3252)	(0.2501)	(0.0496)	(0.0373)	(0.5013)	(0.3735)
Male more educated	0.0488	0.0224	0.2172	0.5776	0.0116	0.0001	0.4938	0.5975
	(0.0564)	(0.0448)	(0.5369)	(0.4083)	(0.0611)	(0.0491)	(0.6235)	(0.5000)
Non-rankable degree	0.0831	0.0986	1.3728	1.6644	0.2916*	0.2564*	3.5910*	2.9083**
	(0.2284)	(0.1482)	(2.0635)	(1.2959)	(0.1482)	(0.0999)	(1.4593)	(0.9985)
Science	0.0574*	0.0727**	0.9701***	0.9189***	0.0444	0.0406	0.5336*	0.7062**
	(0.0281)	(0.0234)	(0.2711)	(0.2158)	(0.0236)	(0.0209)	(0.2476)	(0.2152)
Commerce	0.0558*	0.0535*	0.4692	0.3747	0.0074	0.0618	0.5900	1.2313**
	(0.0279)	(0.0238)	(0.2654)	(0.2190)	(0.0466)	(0.0356)	(0.5229)	(0.3771)
Other field	0.0839	0.0639	0.1661	0.4733	-0.2849	-0.0266	0.6582	1.8935
	(0.0881)	(0.0684)	(0.8389)	(0.6334)	(0.2053)	(0.1164)	(2.3068)	(1.2467)
Calcutta	0.0441*	0.0601***	0.5010*	0.5145***	0.0626*	0.0605**	0.9589**	0.6954**
	(0.0205)	(0.0160)	(0.1957)	(0.1468)	(0.0287)	(0.0232)	(0.3092)	(0.2414)
Same location	0.0715	0.0400	0.2603	0.3765	-0.0179	-0.0207	-0.0462	-0.1084
	(0.0468)	(0.0387)	(0.4501)	(0.3577)	(0.0389)	(0.0331)	(0.4131)	(0.3410)
Same family origin	0.0336	0.0349	0.4720	0.1820	0.0913**	0.0691**	0.5997	0.6442*
	(0.0265)	(0.0218)	(0.2558)	(0.2019)	(0.0309)	(0.0249)	(0.3307)	(0.2602)
Log income	0.1641***	0.1494***	1.3992***	1.2974***				
	(0.0281)	(0.0222)	(0.2655)	(0.2022)				
Log wage	0.0951***	0.0860***	0.8867***	0.8047***				
	(0.0212)	(0.0168)	(0.2037)	(0.1536)				

Continued on next page

	Ads placed by females				Ads placed by males			
	Considered		Rank		Considered		Rank	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Skin tone					-0.0529***	-0.0421***	-0.4603**	-0.5388***
					(0.0143)	(0.0118)	(0.1494)	(0.1209)
Beautiful					0.0151	0.0170	0.4348	0.1823
					(0.0262)	(0.0219)	(0.2757)	(0.2241)
Very beautiful					0.0915	0.0855*	0.4869	0.6153
					(0.0505)	(0.0419)	(0.5124)	(0.4259)
Diff. in quality less than ptile	50th	75th	50th	75th	50th	75th	50th	75th
N	2767	4141	2488	3753	2048	2909	1762	2553

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Standard errors in parentheses. Ads placed by females (males) received letters by males (females): the first four columns refer to decisions made by females regarding prospective grooms, the last four to decisions made by males regarding prospective brides

* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.6: **Responses for letters, top four castes only**

	Ads placed by females			Ads placed by males		
	Considered-OLS	Considered-Logit	Rank	Considered-OLS	Considered-Logit	Rank
	(1)	(2)	(3)	(4)	(5)	(6)
Same caste	0.1636***	0.8372***	1.6650***	0.1047*	0.6521**	0.9490*
	(0.0408)	(0.2017)	(0.3041)	(0.0503)	(0.2180)	(0.4200)
Diff. in caste	-0.0203	-0.0389	-0.2100	0.0307	0.1188	0.6039**
	(0.0157)	(0.0862)	(0.1274)	(0.0204)	(0.0989)	(0.1996)
Same caste*only within	0.2760		4.0097*	0.2206		2.5592
	(0.2504)		(1.6520)	(0.1946)		(1.5047)
Diff. in caste*only within	0.1630		1.5846**	0.0173		-0.2654
	(0.0907)		(0.6090)	(0.0827)		(0.6165)
Same caste*no bar	-0.1214		-1.4500**	-0.0283		-0.4768
	(0.0774)		(0.4943)	(0.0868)		(0.7489)
Diff. in caste*no bar	-0.0013		-0.0133	-0.0526		-0.2027
	(0.0301)		(0.1612)	(0.0347)		(0.2678)
Diff. in age	0.0086	0.1785*	0.0384	0.0424**	0.2239**	0.5249***
	(0.0115)	(0.0824)	(0.0551)	(0.0138)	(0.0783)	(0.0941)
Squared diff. in age	-0.0021**	-0.0237***	-0.0124***	-0.0016	-0.0075	-0.0296***
	(0.0008)	(0.0061)	(0.0034)	(0.0010)	(0.0054)	(0.0064)
Diff. in height	1.7176***	11.5875***	12.8167***	0.4528	9.9158*	6.4163
	(0.4304)	(2.7654)	(2.9819)	(0.5064)	(4.2931)	(3.8687)
Squared diff. in height	-4.7533**	-32.3551***	-36.7084***	-5.5546**	-57.2542***	-69.2712***
	(1.5071)	(9.5394)	(10.5597)	(1.8509)	(16.0106)	(14.5440)
High school	0.0893	-0.3359	0.3344	0.1458	0.6317	2.3437**
	(0.2058)	(1.0614)	(1.0421)	(0.1319)	(0.8511)	(0.7957)
Post-secondary	0.1455	-0.0292	0.9657	1.0020		2.8634
	(0.2204)	(1.1724)	(1.1656)	(0.7954)		(1.7153)
Bachelor's	0.0994	-0.1983	0.9457	0.1373	0.3398	2.8282*
	(0.2228)	(1.1747)	(1.1653)	(0.1754)	(1.0892)	(1.1618)
Master's	0.2457	0.6397	1.7441	0.2074	0.7712	3.9660***
	(0.2286)	(1.2091)	(1.2018)	(0.1799)	(1.1094)	(1.1982)
PhD	0.3103	0.9926	1.9778	0.3754*	2.0243	5.6290***
	(0.2335)	(1.2364)	(1.2347)	(0.1875)	(1.1387)	(1.3764)
Same education	0.0698	0.3108	0.5517*	0.0544	0.2778	0.1380
	(0.0400)	(0.2295)	(0.2502)	(0.0516)	(0.2602)	(0.3726)
Male more educated	0.0683	0.3453	1.1132**	-0.0048	-0.1850	0.2927
	(0.0642)	(0.3564)	(0.3964)	(0.0727)	(0.3859)	(0.5242)
Non-rankable degree	0.2176	0.5038	1.6034	0.3889*	1.8667	3.6022***
	(0.2114)	(1.0908)	(1.0982)	(0.1595)	(0.9668)	(1.0440)
Science	0.1027**	0.6910***	1.1189***	0.0266	0.2026	0.4503
	(0.0339)	(0.1962)	(0.2215)	(0.0320)	(0.1624)	(0.2406)
Commerce	0.0690	0.4884*	0.2930	0.0442	0.2986	0.8302*
	(0.0356)	(0.2064)	(0.2310)	(0.0411)	(0.2131)	(0.3260)
Other field	-0.0211	0.2345	0.1823	0.0806	-0.0493	0.4942
	(0.0953)	(0.5211)	(0.5432)	(0.1210)	(0.7079)	(1.0121)
Calcutta	0.0363	0.2345	0.4769***	0.0472	0.2776	0.6114**
	(0.0224)	(0.1239)	(0.1432)	(0.0318)	(0.1689)	(0.2353)
Same location	0.1162*	0.7043*	0.9203*	-0.0082	-0.0137	-0.1505
	(0.0576)	(0.3370)	(0.3757)	(0.0489)	(0.2607)	(0.3615)

Continued on next page

	Ads placed by females			Ads placed by males		
	Considered-OLS	Considered-Logit	Rank	Considered-OLS	Considered-Logit	Rank
	(1)	(2)	(3)	(4)	(5)	(6)
Same family origin	0.0121	0.1294	0.1625	0.0969**	0.6508***	0.9472***
	(0.0311)	(0.1733)	(0.2085)	(0.0344)	(0.1945)	(0.2728)
Log income	0.1254***	0.2514*	1.0116***			
	(0.0222)	(0.1185)	(0.1564)			
Log wage	0.1176***	0.4247**	0.9331***			
	(0.0235)	(0.1306)	(0.1528)			
Skin tone				-0.0343*	-0.2055*	-0.5198***
				(0.0171)	(0.0927)	(0.1261)
Beautiful				0.0214	0.1621	0.0731
				(0.0313)	(0.1644)	(0.2377)
Very beautiful				0.0472	0.4497	0.5465
				(0.0527)	(0.2594)	(0.3878)
N	2295	2045	2191	3944	1474	3570

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the letter writer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the letter and a dummy for both the letter writer and the ad placer not providing caste, age, height, education, location and family origin. All regressions are weighted to reflect the relative proportions of considered and unconsidered letters received by an ad placer. Standard errors in parentheses. Ads placed by females (males) received letters by males (females): the first three columns refer to decisions made by females regarding prospective grooms, the last three to decisions made by males regarding prospective brides.

* significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.7: Probability of writing to a particular ad

	Ads placed by females				Ads placed by males			
	Ad placer selection		Respondent selection		Ad placer selection		Respondent selection	
	LP	Logit	LP	Logit	LP	Logit	LP	Logit
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Same caste	0.0206***	3.4296***	0.1080***	2.1627***	0.0319***	2.3853***	0.1956***	2.2002***
	(0.0013)	(0.3504)	(0.0022)	(0.0672)	(0.0014)	(0.2043)	(0.0049)	(0.0895)
Diff. in caste*Higher caste male	-0.0013	-1.7058	0.0001	0.0609*	-0.0004	0.2302	0.0236***	0.5106***
	(0.0014)	(1.1849)	(0.0009)	(0.0308)	(0.0013)	(0.3532)	(0.0016)	(0.0353)
Diff. in caste*Lower caste male	-0.0011	-2.0820	-0.0092***	-0.3236***	-0.0020	-0.7402*	0.0014	-0.0809*
	(0.0014)	(1.1721)	(0.0007)	(0.0254)	(0.0012)	(0.3519)	(0.0018)	(0.0380)
Same caste*only within	0.0029	13.0267			-0.0059	14.5443		
	(0.0038)	(770.0985)			(0.0033)	(984.4139)		
Diff. in caste*only within	0.0004	-0.0170			0.0011	0.2650		
	(0.0008)	(368.9421)			(0.0007)	(324.9982)		
Same caste*no bar	-0.0046**	-1.4258***			-0.0010	-0.4298		
	(0.0015)	(0.3972)			(0.0016)	(0.2442)		
Diff. in caste*no bar	-0.0003	-0.1701			0.0007	0.3169**		
	(0.0003)	(0.1420)			(0.0004)	(0.1003)		
Diff. in age	0.0003***	0.2974***	0.0042***	0.4822***	0.0005***	0.4746***	0.0085***	0.6196***
	(0.0001)	(0.0562)	(0.0002)	(0.0158)	(0.0002)	(0.0546)	(0.0005)	(0.0228)
Squared diff. in age	-0.0000***	-0.0234***	-0.0005***	-0.0395***	-0.0000***	-0.0398***	-0.0005***	-0.0484***
	(0.0000)	(0.0043)	(0.0000)	(0.0011)	(0.0000)	(0.0044)	(0.0000)	(0.0017)
Diff. in height	0.0435**	17.6596**	0.3241***	13.3879***	0.0452***	9.7321***	0.3539***	6.0564***
	(0.0167)	(5.9477)	(0.0256)	(1.0314)	(0.0099)	(2.0036)	(0.0413)	(0.8609)
Squared diff. in height	-0.1922***	-75.6526***	-1.2001***	-50.3339***	-0.2013***	-43.4930***	-1.9223***	-32.4783***
	(0.0528)	(20.1851)	(0.0747)	(3.3084)	(0.0414)	(8.3431)	(0.1723)	(3.8381)
High school	0.0013	0.7340	0.0176***	0.4294***	-0.0001	13.1424	-0.0135	-0.1717
	(0.0022)	(0.8006)	(0.0040)	(0.1206)	(0.0029)	(702.6814)	(0.0098)	(0.2239)
Post-secondary	-0.0010	0.2473	-0.0159*	-0.7547**	0.0020	14.0290	0.0117	-0.1526
	(0.0035)	(1.0634)	(0.0065)	(0.2810)	(0.0033)	(702.6813)	(0.0118)	(0.2490)
Bachelor's	-0.0006	0.1855	-0.0115***	-0.2506*	-0.0017	13.2529	-0.0360***	-0.6465**
	(0.0021)	(0.7795)	(0.0035)	(0.1125)	(0.0029)	(702.6813)	(0.0095)	(0.2180)
Master's	0.0024	0.8934	-0.0101*	-0.1507	0.0034	13.9488	-0.0378***	-0.7335**
	(0.0023)	(0.8084)	(0.0039)	(0.1256)	(0.0033)	(702.6813)	(0.0109)	(0.2379)
PhD	-0.0005	0.3537	-0.0151***	-0.1832	0.0048	14.0380	-0.0229*	-0.5667*
	(0.0027)	(0.8864)	(0.0045)	(0.1425)	(0.0035)	(702.6813)	(0.0111)	(0.2423)
Same education	0.0022	0.5264	0.0191***	0.5524***	0.0032*	0.7805**	0.0448***	0.8407***
	(0.0012)	(0.2759)	(0.0019)	(0.0575)	(0.0013)	(0.2434)	(0.0047)	(0.0864)
Male more educated	0.0016	0.4578	0.0014	0.0406	0.0021	0.5918	0.0324***	0.7051***
	(0.0016)	(0.4240)	(0.0030)	(0.0915)	(0.0020)	(0.3213)	(0.0062)	(0.1133)
Non-rankable degree	-0.0031	-13.2632	-0.0242*	-0.5629	-0.0018	13.2663	-0.0534	-0.5984
	(0.0131)	(4420.5696)	(0.0098)	(0.4140)	(0.0049)	(702.6816)	(0.0281)	(0.4275)
Science	0.0004	0.0622	-0.0013	0.0553	0.0022	0.2396	-0.0084	-0.0976
	(0.0008)	(0.1794)	(0.0013)	(0.0395)	(0.0012)	(0.1661)	(0.0055)	(0.0939)
Commerce	0.0009	0.2188	0.0013	0.0450	-0.0015	-0.3376	-0.0186***	-0.2452**
	(0.0012)	(0.2561)	(0.0018)	(0.0539)	(0.0013)	(0.1743)	(0.0055)	(0.0945)
Other field	0.0013	0.0839	-0.0053	-0.0701	0.0085**	1.0443**	-0.0602***	-0.5009
	(0.0035)	(0.7779)	(0.0066)	(0.1701)	(0.0032)	(0.3378)	(0.0178)	(0.2599)
Calcutta	0.0097***	1.7482***	-0.0043	-0.1346	0.0097***	1.1826***	0.0062	0.0029
	(0.0017)	(0.4223)	(0.0038)	(0.1150)	(0.0012)	(0.1721)	(0.0049)	(0.0871)
Same location	-0.0007	0.0442	0.0051	0.2150*	-0.0051	-0.4259	0.0088	0.1428
	(0.0026)	(0.5239)	(0.0029)	(0.0889)	(0.0032)	(0.4468)	(0.0046)	(0.0822)
Same family origin	0.0053***	1.3955***	0.0194***	0.4990***	0.0058***	0.8628***	0.0259***	0.3742***
	(0.0008)	(0.2287)	(0.0012)	(0.0364)	(0.0009)	(0.1545)	(0.0027)	(0.0463)
Log income					0.0024**	0.2556*	0.0044	-0.0708
					(0.0009)	(0.1187)	(0.0037)	(0.0683)
Log wage					0.0041***	0.8576***	0.0010	0.0260
					(0.0005)	(0.1070)	(0.0020)	(0.0352)
Skin tone	-0.0012**	-0.3719**	-0.0033***	-0.0927***				
	(0.0004)	(0.1179)	(0.0007)	(0.0219)				
Beautiful	-0.0011	-0.2338	0.0016	0.0264				
	(0.0007)	(0.1671)	(0.0012)	(0.0369)				
Very beautiful	0.0008	0.0304	0.0047	0.0523				
	(0.0015)	(0.3025)	(0.0024)	(0.0683)				
N	49025	49025	147546	144543	70337	69617	53043	52407

All regressions include dummies for caste, for being from West Bengal, dummies indicating non-response for each characteristics, age/height of the respondent/ad placer if no age/height was provided by the ad, age/height of the ad placer if no age/height was provided by the respondent/ad placer and a dummy for both individuals not providing caste, age, height, education, location and family origin. Ads placed by females (males) received letters by males (females): the first four columns refer to decisions made by males regarding which ad placed by females they should write to, the last four to decisions made by females regarding which ads placed by males they should contact. Standard errors in parentheses. * significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.8: **Number of responses received to an ad**

	Ads placed by females		Ads placed by males	
	OLS	Poisson	OLS	Poisson
	(1)	(2)	(3)	(4)
Baidya	0.0199	1.4363	-0.4018***	-32.5365
	(0.0554)	(4.5688)	(0.0387)	(22.6938)
Kshatriya	-0.3880***	-6.4094	-0.4774***	-32.4609
	(0.1017)	(7.0018)	(0.0746)	(38.5897)
Kayastha	0.1941***	4.8539*	0.1565***	14.8425
	(0.0242)	(2.2215)	(0.0176)	(12.0916)
Baisya	-0.2298***	-4.2818	-0.0679**	-6.3319
	(0.0313)	(2.5611)	(0.0214)	(13.7648)
Sagdope	-0.0900*	-2.0499	-0.0344	-3.5924
	(0.0360)	(3.2275)	(0.0253)	(15.8213)
Other non-scheduled castes	-0.5491***	-8.1897	-0.6427***	-28.3260
	(0.1107)	(7.2236)	(0.0673)	(30.0856)
Scheduled castes	-0.0659	-1.2732	-0.5098***	-39.0446
	(0.0670)	(5.5995)	(0.0421)	(23.3959)
Age	-0.0401***	-0.8096**	0.0119***	0.8895
	(0.0031)	(0.2490)	(0.0016)	(1.0717)
Height	1.5551***	35.4319	-0.4142***	-17.6774
	(0.2196)	(19.5507)	(0.1239)	(79.5235)
High school	-0.1107	-1.8582	0.8501***	19.0770
	(0.0761)	(6.5589)	(0.1762)	(55.5553)
Post-secondary	-0.4580	-10.6578	1.6886***	82.9122
	(0.2403)	(20.2488)	(0.1781)	(61.3144)
Bachelor's	-0.0769	-1.2923	1.5513***	67.2765
	(0.0774)	(6.7409)	(0.1756)	(56.9136)
Master's	-0.1423	-2.8572	1.8182***	89.1902
	(0.0808)	(7.0390)	(0.1768)	(58.7970)
PhD/Professional degrees	-0.2741**	-5.4127	1.7035***	77.3746
	(0.0926)	(7.8143)	(0.1767)	(58.3160)
Non-rankable degree	-1.0200***	-14.9420	1.2666***	40.0588
	(0.1777)	(10.7632)	(0.1896)	(69.6573)
Science	0.0463	1.2457	0.2546***	22.4205
	(0.0253)	(2.2666)	(0.0421)	(26.3598)
Commerce	-0.0520	-1.1006	-0.0265	-1.1862
	(0.0346)	(3.0170)	(0.0433)	(26.8366)
Other field	-0.6742*	-5.9297		
	(0.2846)	(14.3313)		
Calcutta	0.4087***	8.6102	0.1608***	20.7122
	(0.0684)	(5.3780)	(0.0164)	(13.4021)
From West Bengal	0.1941***	4.6963*	0.4275***	29.7894
	(0.0228)	(2.0787)	(0.0271)	(15.4041)
Log income			-0.2129***	-16.0723
			(0.0180)	(11.4682)
Log wage			0.0190	3.6086
			(0.0200)	(13.2790)
Skin tone	-0.2570***	-5.1665***		
	(0.0166)	(1.2562)		
Very beautiful	0.2804***	9.0867*		
	(0.0369)	(3.8408)		
Beautiful	0.0147	0.3033		
	(0.0243)	(2.1623)		
N	5788	5788	4075	4075

Standard errors in parentheses. All regressions include dummies indicating non-response for each characteristics. *significant at 5%; ** significant at 1%; *** significant at 0.1%

Table C.9: **Couples' characteristics, variances of the algorithm**

	Women propose		Balanced sex ratio	
	2.5 ptile	97.5 ptile	2.5 ptile	97.5 ptile
	(1)	(2)	(3)	(4)
Age difference	**5.4765**	**6.4272**	4.5947	5.3435
Age correlations	**0.8079**	**0.9376**	**0.7370**	**0.8997**
Height difference	**0.1049**	**0.1222**	**0.1128**	**0.1297**
Height correlations	0.7752	0.8955	0.7536	0.8742
Same caste	0.8439	0.9556	0.8598	0.9631
Caste difference	*0.1111*	*0.6316*	**-0.0743**	**0.1620**
Caste correlation	**0.5680**	**0.9296**	**0.5714**	**0.9756**
Same education level	**0.2090**	**0.8019**	**0.3248**	**0.7812**
Education difference	-0.5250	-0.0098	**-0.0656**	**0.4133**
Education correlations	**0.2591**	**0.6586**	**0.3659**	**0.7289**
Same family origin	0.9893	1.0000	0.9579	1.0000
Family origin difference	**-0.0067**	**0.0064**	**-0.0064**	**0.0347**
Family origin correlations	0.9766	1.0000	0.9079	1.0000
Same residence	**0.0000**	**1.0000**	**0.0000**	**1.0000**
Location correlations	**-0.7986**	**1.0000**	**-0.8419**	**1.0000**
Log wage difference	*-0.3380*	*0.0815*	-0.4980	-0.0539
Log wage correlations	**-0.2233**	**0.3461**	**-0.1700**	**0.3497**
Income difference	**-491999.30**	**40416.89**	**-0.02**	**14500.29**
Income correlations	**-1.0000**	**1.0000**	**-1.0000**	**1.0000**
Quality difference	0.1566	0.1758	0.1662	0.1887
Quality correlation	**0.0785**	**0.4057**	*0.2705*	*0.5355*

Entries in bold correspond to characteristics where the observed characteristics fall within the estimated confidence interval. Entries in italic have overlapping confidence intervals with the observed distribution.

Figure C.1: Correlations between coefficients of the considered and rank regressions, ads placed by females

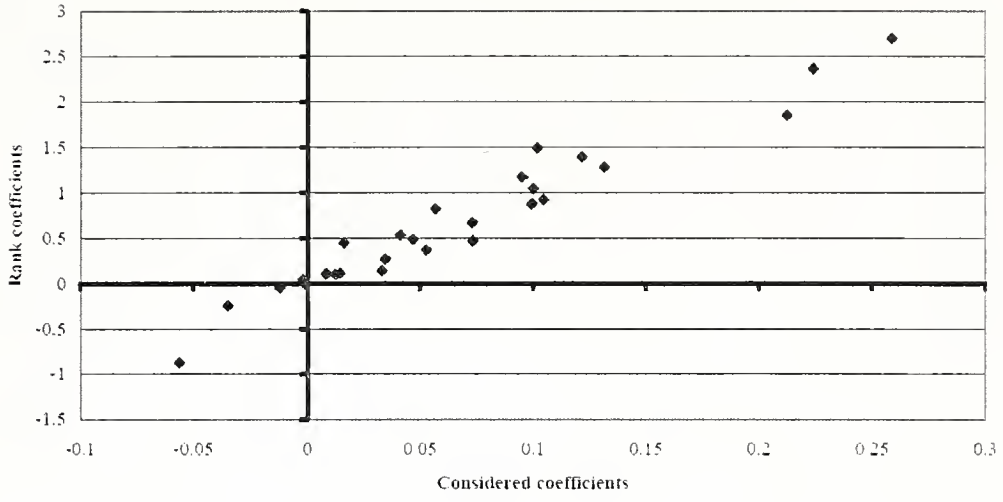


Figure C.2: Correlations between coefficients of the considered and rank regressions, ads placed by males